Turbulence Modeling of Strongly-Coupled Particle-Laden Flows

by

Sarah Beetham

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
(Mechanical Engineering and Scientific Computing)
in The University of Michigan
2021

Doctoral Committee:

Assistant Professor Jesse Capecelatro, Chair
Professor Kevin Maki
Assistant Professor Aaron Towne
Professor Angela Violi

Sarah Beetham

snverner@umich.edu

ORCID iD: 0000-0002-2823-2394

for Jack and Katie

ACKNOWLEDGEMENTS

There is an old proverb, "It takes a village to raise a child." I would argue that this sentiment can also be extended to PhD students. Here, I'd like to thank a number of people, who I consider to be the 'village' that raised me over the last five years.

First, I cannot over express my gratitude to my advisor, Professor Jesse Capecelatro, for adopting me into his lab as a second year graduate student. I often recall that period of my academic journey and firmly believe that had he not taken me under his advisement, I would have likely not finished my studies. Under his supervision and with his encouragement, I've been able to grow as a researcher and pursue many areas of my own interests. I hope to apply the many lessons I learned as his student to my own future career advising and mentoring others.

Next, it goes without saying that I would have not been able to have accomplished any of the work in this dissertation without the unwavering support of my family. My husband, Tommy, two children, Jack and Katie, and our dogs, Marty and Casey, always helped me to maintain a balance and keep even the most critical Reviewer 2 in perspective. I know the stress I carried over the last five years was difficult for more than just me, and I deeply appreciate their patience and love.

I also cannot adequately describe the ways in which my parents have supported and encouraged me on this journey, beginning in middle school when they fostered my love for math and science and stood up to teachers who questioned 'where would she ever have a use for algebra.' They enabled me to be able to be on campus, to travel to conferences and always were willing to listen when I needed to unload stress.

Over the years, my parents have always supported the pursuit of my dreams, and I can only hope to give even a fraction of that level of support to my own children as they grow.

My mother-in-law, Patti, has also been a tremendous and consistent form of support and encouragement. She drove weekly from Ontario to stay with Jack and Katie so I could go to campus and without her unwavering commitment, I would have not been able to be writing this dissertation now.

To my dear friend, Linda Schilbe, I also owe a great deal of gratitude. She also helped watch Jack while I studied and constantly kept me in prayer.

Finally, I'd like to thank my lab-village and dear friends, Greg Shallcross, Aaron Lattanzi, Yuan Yao, Ali Kord and Meet Patel. You have all been a constant source of encouragement and advice (and at times comic relief) and I look forward to referring to you all as my academic brothers forever.

The material presented in this dissertation is supported by the National Science Foundation (NSF CAREER, CBET-1846054) and in particular the NSF Graduate Research Fellowship. I owe a debt of gratitude to the NSF GRFP, which allowed me the ability to shift my research focus entirely after my first year of graduate school. Without this source of funding and the flexibility that it provided, following my passion would have been a more tumultuous endeavor.

Finally, the computing resources and assistance provided by the staff of Advanced Research Computing at the University of Michigan, Ann Arbor is greatly appreciated. Additionally, this work used the Extreme Science and Engineering Discovery Environment (XSEDE), which is supported by National Science Foundation grant number ACI-1548562 [1].

TABLE OF CONTENTS

LIST OF FIGURES

Figure

LIST OF TABLES

Table

LIST OF APPENDICES

Appendix

ABSTRACT

Turbulent, disperse two-phase flows are pervasive in nature and industry. Some contemporary examples include the dispersion of respiratory droplets when coughing or sneezing, sedimentation transport in rivers and the upgrading of agricultural waste into usable biofuels. In many systems, the disperse phase (e.g., solid particles, liquid droplets, gas bubbles) modifies the turbulence in the carrier phase, giving rise to complicated flow features such as dense clusters (or bubble clouds) and regions nearly void of particles. This heterogeneity predicates a wide range of length- and time-scales, making fully-resolved computations at scales of interest intractable, even on modern super computers. Thus, the Reynolds-Averaged Navier–Stokes (RANS) equations, which depend heavily upon modeling, continue to be the primary tool for large-scale computations of both single and multiphase turbulence.

Despite the prevalence and importance of multiphase systems, developing accurate models, especially for the multiphase RANS equations, has remained an open challenge. Due to the large parameter space, brute-force modeling approaches are infeasible. Further, the presence of a disperse phase can generate energy at the small scales (i.e., wakes past particles) which induces turbulence at large scales. This is directly in conflict with energy cascade theory from single-phase turbulence, thus making extensions from traditional single-phase turbulence modeling inadequate.

Due to the lack of accurate, tractable models for the multiphase RANS equations, researchers and practitioners must rely on closures that make idealistic simplifications such as uniformity in the disperse phase or perfect mixing. These sweeping assumptions lead to large errors in predicting quantities of interest (like the thermal entrance

length or rate of thermochemical conversion), because important multiphase physics have been neglected. To demonstrate this shortcoming, the conversion of biomass (woody, agricultural wast) into biofuel is simulated using highly-resolved, Eulerian–Lagrangian simulations. In this example, highly-resolved data is compared with predictions of an idealized model typical of industry that assumes uniform particles and perfect mixing. It is found that ignoring heterogeneity in the system results in an *under prediction* of biofuel yield by 33%. This underscores why neglecting multiphase physics presents a principal challenge for upscaling reactors to industrial scales.

Motivated by this disparity, the main objective of this work is to develop a modeling framework capable of accurately translating highly-resolved data into models that are rooted in knowledge of physics, interpretable and easy to share within and across communities. While this framework will benefit specific multiphase applications, such as fluidized bed reactors, it also has broader implications as the framework itself presents a generalized means to model quantities of any dimension, from scalars to tensors, with guaranteed invariance in compact, algebraic form.

In this work, two modeling methodologies are developed for the first time: (1) Sparse regression with embedded invariance and (2) Sparse regression blended with gene expression programming (GEP). These methodologies are validated and tested on several canonical *single*-phase turbulent flows and extended to gravity-driven gas–solid flows of increasing dimensionality. Here, a minimal invariant tensor basis is derived for the first time for this class of flows and new models are proposed. Finally, new scaling and models are developed for systems with spatially evolving heat transfer.

The modeling advancements set forth in this work are important for enabling industrial capability for predicting complex, gas-solid flows. In the context of the conversion of biomass to biofuel, improved reduced order models are needed to enable the scale up of circulating fluidized bed reactors from lab-scale to industrial-scale.

This scale up is essential for harnessing the carbon neutral, high efficiency benefits of this renewable energy source and meeting climate and emissions targets.

CHAPTER I

Introduction

1.1 The ubiquity of disperse, two-phase flows

The focus of this dissertation is modeling disperse, two-phase flows. Here, the term 'two-phase' refers to a wide range of states, including: liquid droplets sprayed in a gas, solid particles suspended in a gas or liquid, and bubbles dispersed in a liquid. In many of these flows, the volume fraction of the disperse phase is large enough such that the effect the disperse phase has on the carrier phases is increasingly substantial (see, e.g., Figure 1.4). In other words, the coupling of the phases causes the carrier phase to behave entirely differently in the presence of a disperse phase as compared to its behavior as a single phase. These types of flows are pervasive in many important contexts within nature, the human body and industry.

Some examples in the environment (see Fig. 1.1) include sedimentation transport in rivers and oceans and the flow of groundwater or harmful contaminants through porous media (such as the ground). These systems are of great societal importance and our ability to characterize their behavior is critical for improving the health and safety of communities. Recently, understanding and mitigating the mechanisms related to dispersion of respiratory droplets into the air due to coughing and sneezing has become of utmost importance for controlling the spread of viruses like COVID-19 and informing policy decisions and public health guidelines. Additionally, develop-

Figure 1.1: Flows containing a disperse phase are pervasive in nature and industry.

ing clean, efficient energy sources has also become an urgent societal need. Nearly all chemical transformation processes, such as the upgrading of woody, agricultural waste (termed biomass) into biofuel, require the use of two-phase reactors. Many laboratory-scale systems have demonstrated great promise, but the scale-up to commercial application remains an open challenge, in no small part due to the complexity associated with the collisional, two-phase nature of the system. For each of these applications, understanding and predicting flow behavior is imperative for achieving improved system designs, better physical understanding or control strategies.

1.2 The challenge of modeling turbulent, two-phase flows

Turbulent flows, on their own, are historically challenging, even without the addition of a disperse phase. Thus, before discussing the complexities of *disperse, two-phase* turbulence, it is instructive to highlight some key aspects of single-phase turbulence.

1.2.1 Single phase turbulence

At the most fundamental level, turbulence is characterized by velocity field fluctuations, denoted as u_i', from the mean, $\langle u_i \rangle$, where u_i is the velocity vector in index notation and angled brackets denote a Reynolds average. This yields the Reynolds decomposition [14] for the velocity field, given as

$$u_i = \langle u_i \rangle + u_i' \tag{1.1}$$

This decomposition enabled the first step toward understanding turbulent flows: derivation of the Reynolds-Averaged Navier–Stokes (RANS) equations [14]. In this formulation, the incompressible Navier–Stokes equations,

$$\frac{\partial u_i}{\partial t} + u_j \frac{\partial u_i}{\partial x_j} = \frac{\partial}{\partial x_j} \left[-\frac{1}{\rho} p \delta_{ij} + \nu \left(\frac{\partial u_i}{\partial x_j} + \frac{\partial u_j}{\partial x_i} \right) \right] + g_i, \tag{1.2}$$

are averaged over homogeneous dimensions, giving rise to a set of mean equations referred to the Reynolds-averaged Navier–Stokes (RANS) equations [14]. The RANS equations are written as

$$\frac{\partial \langle u_i \rangle}{\partial t} + \langle u_j \rangle \frac{\partial \langle u_i \rangle}{\partial x_j} = \frac{\partial}{\partial x_j} \left[-\langle p \rangle \delta_{ij} + \nu \left(\frac{\partial \langle u_i \rangle}{\partial x_j} + \frac{\partial \langle u_j \rangle}{\partial x_i} \right) - \langle u_i' u_j' \rangle \right] + g_i, \tag{1.3}$$

where p is the pressure, ν is the kinematic viscosity, δ_{ij} is the Kroneker delta and $\langle u_i' u_j' \rangle$ is the Reynolds stress tensor. The Reynolds stress tensor represents the cross correlation in velocity fluctuations and its divergence yields a quantity with units of force per volume. This can be thought of as the force a volume of fluid feels due to the turbulence in the flow, and is null for laminar flow.

When the RANS equations are non-dimensionalized with a characteristic velocity, U, and length scale, L, two important groupings arise: the Reynolds number (Re)

and the Froude number (Fr):

$$\frac{\partial\langle\hat{u}_i\rangle}{\partial\hat{t}} + \langle\hat{u}_j\rangle\frac{\partial\langle\hat{u}_i\rangle}{\partial\hat{x}_j} = \frac{\partial}{\partial\hat{x}_j}\left[-\langle\hat{p}\rangle\delta_{ij} + \frac{1}{\mathrm{Re}}\left(\frac{\partial\langle\hat{u}_i\rangle}{\partial\hat{x}_j} + \frac{\partial\langle\hat{u}_j\rangle}{\partial\hat{x}_i}\right) - \langle\hat{u}_i'\hat{u}_j'\rangle\right] + \frac{1}{\mathrm{Fr}^2}\hat{g}_i. \quad (1.4)$$

These non-dimensional groups are defined as,

$$\mathrm{Re} = \frac{UL}{\nu} \qquad \text{and} \qquad \mathrm{Fr} = \frac{U}{\sqrt{gL}}. \quad (1.5)$$

Here, the Reynolds number is a ratio of inertial to viscous forces and the Froude number is the ratio of inertial to gravitational (or other external body) forces. Since the inverse of the Reynolds number multiplies the viscous term, this implies that when Reynolds number is large, turbulence dominates and when it is small, viscous effects are more important. For this reason, Re is a classical indicator of the level of turbulence in a flow and the range of scales present.

All the terms appearing in the RANS equations are expressed in terms of mean flow quantities, except for the Reynolds Stresses, which requires modeling. The identification of suitable models for this term has been coined 'the closure problem,' since the RANS equations cannot be solved without a model, or closure, for this tensor.

1.2.2 A brief history of single-phase turbulence modeling

The pursuit of accurate closures of the Reynolds stress has been the subject of scientific inquiry for more than a century and is briefly summarized here and in Fig. 1.2.

Shortly after Reynolds' derivation of the RANS equations, the Boussinesq approximation [15] was postulated. This hypothesis still endures today and employs the notion that the Reynolds stresses can be related to the mean shear tensor by way

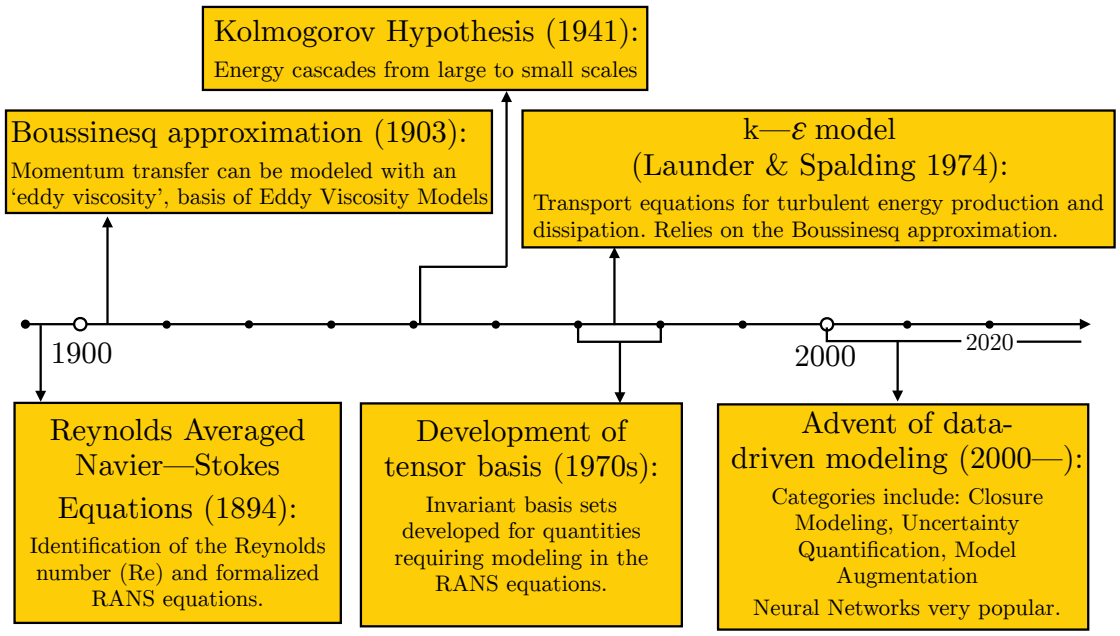


Figure 1.2: A brief overview of single-phase turbulence modeling.

of a turbulent viscosity, ν_t, given as

$$\langle u_i' u_j' \rangle = -\nu_t \left(\frac{\partial \langle u_i \rangle}{\partial x_j} + \frac{\partial \langle u_j \rangle}{\partial x_i} \right) - \frac{2}{3} k \delta_{ij}, \tag{1.6}$$

where k denotes the turbulent kinetic energy, defined as half the trace of the Reynolds stress tensor, $k = \frac{1}{2}\langle u_i' u_i' \rangle$. This enables the Reynolds stress term to be combined with the viscous diffusion term in (1.3). In other words, the eddy viscosity augments the kinematic viscosity in the RANS equations to account for turbulence. This theory is the foundation of the numerous eddy viscosity models that have been proposed over the last century.

Subsequently, Kolmogorov penned the 'K41 theory' [2] which identified the flow of turbulent energy from large to small scales. Here, energy is generated at large scales (e.g., large eddies generated by vortex shedding) and is cascaded through an inertial range after which it is dissipated to heat at the small scales due to viscous diffusion. In the inertial range (the region between the low wave number, large scale regime

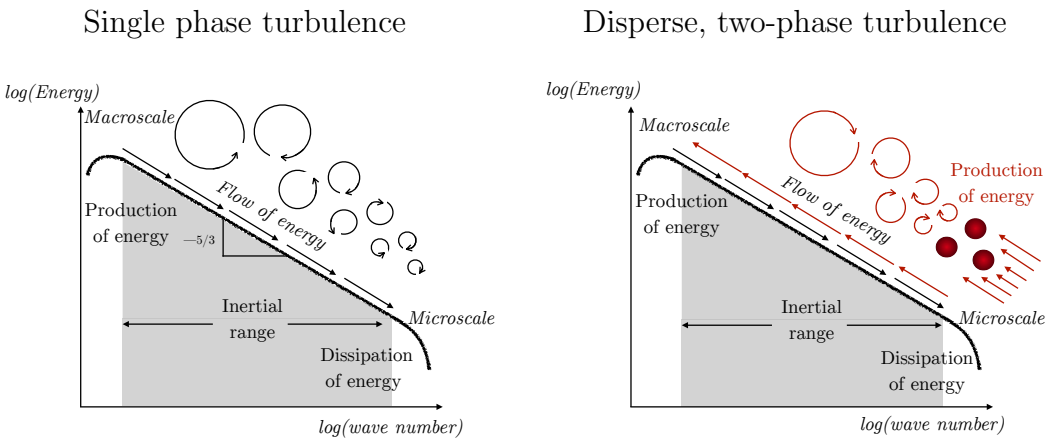


Figure 1.3: The energy cascade from K41 theory [2] (left) breaks down for disperse, two-phase flows (right), where turbulent energy can be generated at the microscale and cascade *up* to the macroscale.

that generates turbulent energy and the high wave number, small scale regime that dissipates it), turbulent energy is transported from high to low wave numbers and this transfer scales as $-5/3$ (see Figure 1.6).

This theory also gave rise to the definition of the Kolmogorov length scale, η. This quantity is the smallest turbulent length scale and is defined as

$$\eta = \left(\frac{\nu^3}{\varepsilon}\right)^{1/4}, \tag{1.7}$$

where ν is the kinematic viscosity and ε is the dissipation rate of turbulent kinetic energy. At scales smaller than this quantity, viscous effects dominate and turbulent energy is dissipated to heat. The Kolmogorov length scale and its associated time scale, are critical for ensuring that simulations of turbulent flow are properly resolved.

Several decades later, two simultaneous developments occurred: (1) the development of a minimal invariant tensor basis and (2) the formulation of the k-ε equations.

The former was developed in the context of theoretical linear algebra, but ultimately had implications for turbulence modeling. In this area of work [see, e.g., 16, 17], proofs were written that established that any tensor, $\mathcal{D}_{ij}$ may be exactly

described by an infinite sum of coefficients, $\beta^{(n)}$ and a tensor basis, $\mathcal{T}_{ij}^{(n)}$:

$$\mathcal{D}_{ij} = \sum_{n=1}^{\infty} \beta^{(n)} \mathcal{T}_{ij}^{(n)}, \tag{1.8}$$

where the coefficients may be nonlinearly dependent upon the principal invariants of the basis tensors. Using arguments of the Cayley Hamilton theorem, special cases exist for which this infinite sum can be reduced to a *finite* sum, thus yielding a minimal invariant basis.

Given that many of the quantities of interest in turbulence (in particular the Reynolds stresses) are tensors, these theorems have proven to have important utility in the context of turbulence modeling. One important utility is the fact that the basis tensors and coefficients are related *linearly*, which is critical for imposing form or Galilean invariance (see Chapter III for additional details).

At the same time, Launder et al. [18] formulated transport equations for the turbulent kinetic energy production (k) and dissipation (ε),

$$\frac{\partial k}{\partial t} + \langle u_i \rangle \frac{\partial k}{\partial x_i} = \frac{\partial}{\partial x_i}\left(\frac{\nu_t}{\sigma_k}\frac{\partial k}{\partial x_i}\right) + P_k - \varepsilon \tag{1.9}$$

$$\frac{\partial \varepsilon}{\partial t} + \langle u_i \rangle \frac{\partial \varepsilon}{\partial x_i} = \frac{\partial}{\partial x_i}\left(\frac{\nu_t}{\sigma_\varepsilon}\frac{\partial \varepsilon}{\partial x_i}\right) + C_{\varepsilon 1}\frac{\varepsilon}{k}P_k - C_{\varepsilon 2}\frac{\varepsilon^2}{k} \tag{1.10}$$

$$P_k = \nu_t \left(\frac{\partial \langle u_i \rangle}{\partial x_j} + \frac{\partial \langle u_j \rangle}{\partial x_i}\right)\frac{\partial \langle u_i \rangle}{\partial x_j}, \tag{1.11}$$

where $C_{\varepsilon 1}$, $C_{\varepsilon 2}$, σ_ε and σ_k are constants and ν_t is the turbulent viscosity as in (1.6). These equations, in combination with the Boussinesq approximation, provide a construct for closure of the RANS equations and are still used widely today. The turbulent viscosity itself is modeled in a number of ways, including the Prandtl mixing length [19] ($\nu_t = |\frac{\partial u}{\partial y}| l_m^2$, where l_m is the mixing length) and the Smagorinsky model [20].

After this period of advancement through the mid-century, traditional turbulence

modeling experienced a period of latency, where progress was primarily incremental. In the early 21st century, the 'age of data' [21] dawned and large-scale, highly resolved simulations of turbulent flow became more widely accessible. With this new wealth of data, the turbulence modeling community turned from more traditional approaches to data-driven modeling. This resulted in an explosion of new interest and productivity in turbulence modeling.

1.2.3 Data-driven efforts in turbulence modeling

The flurry of activity in recent years has including uncertainty prediction and quantification, model calibration and augmentation and the generation of entirely new models. Several recent works have utilized machine learning (neural networks are particularly popular [22, 23, 24, 25, 26, 27, 28, 29, 30]) to translate large amounts of experimental or computational data into turbulence models. Neural networks have shown relatively exceptional performance outside the region in which they were trained. However, as a departure from traditional modeling techniques, these methods are inserted modularly, as a 'black box,' into an existing flow solver. Thus, while they have displayed a high level of performance on a wide range of turbulent flows, the closure does not satisfy the interpretability condition necessary for making physical inferences.

Several studies have taken alternate approaches to NNs and instead use symbolic methods to arrive at closed form, algebraic models. Gene Expression Programming [31, 32, 33, 34, 35] and random forest regression [36, 37] have become increasingly popular methodologies. Additionally, Brunton et al. [38] developed a strategy based on sparse regression that identifies the underlying functional form of the nonlinear physics by optimizing a coefficient matrix that acts upon a matrix of trial functions. While this method requires knowledge about the physics of the system under configuration (in order to make informed selections of the trial functions), it can be

reasonably assumed that the modeler is not entirely naive. In fact, traditional modeling techniques have relied nearly exclusively on this notion.

A large number of data-driven approaches implemented in turbulence modeling augment or correct existing models. While this approach is acceptable for simple, single-phase turbulence (e.g., flows in equilibrium that are free of strong adverse pressure gradients), it is insufficient for more complex turbulent flows, such as disperse two-phase flows [39, 40, 41, 42, 6] or turbulent combustion [43, 44], in which the fundamental assumption of an energy cascade breaks down due to production at the smallest scales (see Fig. 1.3). In these cases, existing closures adopted from single-phase flows are not appropriate, which precludes an augmentation modeling approach.

1.2.4 Turbulent particle-laden flows

In gas-solid flows, the class of flows considered in this work, the addition of a disperse phase substantially complicates flow physics as compared with single-phase turbulence. For a given disperse phase with particles of diameter, d_p, density, ρ_p and volume fraction ε_p, several dimensionless groups arise in addition to those appearing for single-phase flow. These include the density ratio between the phases, ρ_p/ρ_f, the mass loading defined as $\varphi = (\rho_p\langle\varepsilon_p\rangle) \,/\, (\rho_f\varepsilon_f)$, the mean volume fraction, $\langle\varepsilon_p\rangle$ and the ratio between particle diameter and the Kolmogorov length scale, d_p/η. Additionally, a Stokes number can be defined as $\mathrm{St} = \tau_p/\tau_\eta$, where the Stokes relaxation time is $\tau_p = \rho_p d_p^2/(18\mu_g)$ and $\tau_e ta$ is the Kolmogorov time scale.

Depending on the values of these parameters, a wide array of regimes are possible. As shown in Figure 1.4, when the solid phase volume fraction is small, particles act as tracers and have a minimal effect on the fluid. This is referred to as a one-way coupled flow (e.g. the fluid affects the particles but the particles have a negligible effect on the fluid). As the volume fraction increases, particles have a greater effect on modulating turbulence in the carrier phase, referred to as two-way coupling. Here,

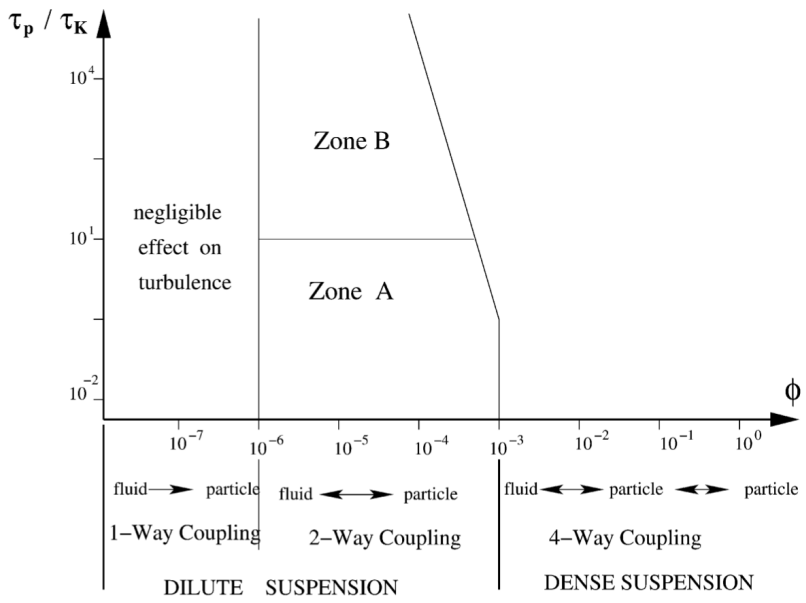


Figure 1.4: Regime map for multiphase flows [3]. Here, τ_p/τ_K denotes the ratio of the Stokes particle response time and the Kolmogorov time scale.

two-way coupling refers to the fluid affecting the particles and particles affecting the fluid. This coupling is particularly important for flows that have a mass loading, defined as $(\rho_p\langle\varepsilon_p\rangle)\rangle/(\rho_f\langle\varepsilon_f\rangle)$, greater than unity and particle diameter larger than the Kolmogorov length scale. Finally, in the case of very dense particle suspensions, in addition to having an important effect on fluid phase turbulence, inter-particle collisions are also important. This is referred to as four-way coupled flow.

The particular focus of this work is on moderately dense gas-solid flows that fall in the four-way coupled regime. In this regime, interphase coupling is important due to the large density ratio and moderate volume fraction and mass loading and the coupling between the disperse and carrier phases generates a wide range of complex phenomenon beyond what is observed in single-phase turbulence. Some of these effects include the formation of wakes past individual particles, chemical adsorption/absorption and convection/conduction at the surface of a particle, and the formation of highly organized structures like particle clusters (see Fig. 1.5). Each of these behaviors give rise to large-scale heterogeneity in the flow that can either

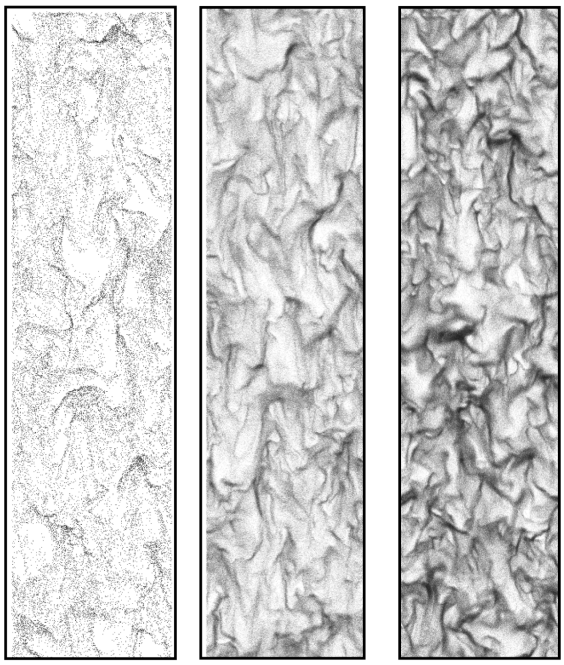


Figure 1.5: When a random suspension of particles at sufficient mass loading settle under the influence of gravity, coherent structures form, seen as clusters here. Three mean volume fractions of particles are shown, 0.1%, 2.55% and 5%, from left to right. More details can be found in Chapter IV.

augment or hinder large-scale processes like mixing or catalytic conversion. In other words, these behaviors generate turbulent energy at the *small* scales which then flows to large scales. This is in direct contradiction to classical single phase theory (see Figure. 1.3).

To illustrate the importance of these interactions and the separation of scales present, consider the conversion of biomass to biofuel in a circulating fluidized bed reactor [6] (see also, Figure 1.6). In this system, cool solid biomass particles and hot sand particles are fluidized with hot nitrogen gas in a tall duct or tube. This process allows for the rapid heating of the biomass particles, causing them to pyrolyze (i.e., breakdown into usable biofuel and other products). At the particle scale, the physics at play include gas-phase convection between particles, conduction at the particle surface, wakes shed by particles and catalytic reactions in which mass is transferred from the solid biomass particles to the gaseous carrier phase. Each of these influence

mesoscale structures in the form of clusters (see Fig. 1.5) and voids that introduce heterogeneity (e.g., hot and cold spots) throughout the flow. This process effectively 'demixes' the carrier flow, which reduces the contact between phases and degrades biomass conversion efficiency, thereby driving up cost (this system is explored in more detail in Chapter II). Thus, understanding and predicting these types of flows is of utmost importance.

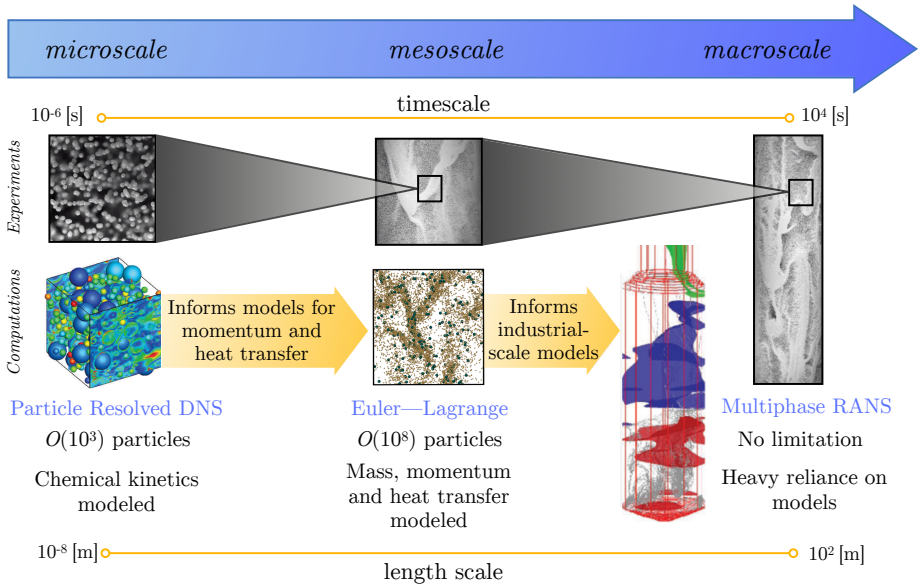


Figure 1.6: Computations capable of capturing reactive, multiphase flows (such as biomass pyrolysis in a circulating fluidized bed reactor) rely on a wide range of modeling across scales. Because of the spectrum of flow conditions and behavior that exist across scales, it is not appropriate to inform RANS models using DNS data. Top: High-speed imaging of particles in a circulating fluidized bed reactor [4], Bottom: Modeling strategies, PR-DNS and Multiphase RANS simulation images from [5] and Euler–Lagrange simulation from [6].

1.2.5 Modeling disperse two-phase flows

While physical experiments can be informative and useful for studying disperse, two-phase flows, computations are often more attractive from an economic and timing standpoint and also allow more invasive probing of the physics at play without affecting fundamental behavior. However, capturing these complex physics is numer-

ically challenging. Due to the two-way coupling between the phases present in these flows, an enormous span of length– and time–scales are often present. Processes that exist at the particle-scale entail extremely small length– and time–scales and the full system spans very large scales (see Figure 1.6).

The most comprehensive strategy for simulating disperse, two-phase flows is particle resolved direct numerical simulation (PR-DNS) [5], in which the computational grid spacing is much smaller than the particle diameter (i.e. tens to hundreds of times smaller) in order to resolve the hydrodynamics and heat transfer. While this technique has minimal reliance on modeling, it is limited to $\mathcal{O}(10^3)$ particles. While unable to capture larger scale phenomenon (like clustering), these simulations are useful for informing models for mass, momentum and heat transfer for coarse-grained simulations based, such as Euler-Euler and Euler-Lagrange methods. These approaches are based on a local volume averaging (or filtering) of the governing equations, and require models for processes occurring at the particle scale (e.g., drag, convection, etc.). Thus, these methods are analogous to large eddy simulation (LES), in the sense that they separate large and small scales, requiring models for the unresolved small scales at the particle surface. While they make larger systems accessible (on the order of hundreds of millions of particles), they are still computationally expensive and further, industrial-scale systems typically contain billions of particles. This has given rise to the prevalence of Reynolds-average Navier–Stokes (RANS) approaches, which rely heavily on models for both large– and small–scale behavior, especially for industrial applications.

Many systems of societal importance are designed using overly simplified models in the RANS equations, leading to inefficiencies and lost cost. These models, while poor, are often the only choice engineers have, as accurate models either do not exist or are not tractable (even on modern super computers) at scales of interest. The difficulty in developing improved models stems from the fact that these systems are

both turbulent *and* multiphase. Predicting and describing turbulence, by itself, has been an open challenge in engineering for the last century and the addition of a second, dispersed phase makes the problem of closure even more challenging.

1.2.5.1 An even briefer history two-phase turbulence modeling

The history of two-phase turbulence modeling is substantially shorter than its single-phase counterpart and largely begins in the late 1980s. During this period and up until the present, two-phase turbulence models have primarily been formulated by drawing upon extensions to single-phase turbulence models (see e.g. [45], [46], [47], [48], [49], [50] [51] [52] [53]). In an attempt to make connections with single-phase theory, these models were derived directly from the Navier–Stokes equations. Due to the fundamental difference in behavior between single and multiphase turbulence, these approaches can only attain some degree of accuracy and this is limited to very dilute disperse phases (i.e. one-way coupled regimes) where the departure from single-phase theory is less pronounced (see Figure 1.4).

In contrast to modeling by analogy with single-phase flow, Fox [39] developed the exact Reynolds-averaged equations for collisional, fluid–particle flows. There, it was demonstrated that directly averaging the Navier–Stokes equations fails to capture important two-phase interactions. Instead, it was shown that phase averaging the mesoscale (locally averaged) equations results in a set of equations that explicitly account for two-way coupling contributions.

Following the work of Fox [39], the RANS equations for collisional, fluid-particle flows have been rigorously derived by Capecelatro et al. [41], and although they are exact, they lead to a number of unclosed terms that require modeling (this is the focus of Chapter IV). Because of the coupling between phases, nearly all of these terms do not appear in the single-phase RANS equations, making extensions to single-phase turbulence models inadequate, particularly outside the dilute regime

where the presence of particles have a more pronounced effect. Further, the larger parameter space as compared with single phase flow (as previously discussed) makes a brute-force modeling approach challenging if not intractable, predicating the need for a data-driven modeling technique instead.

Despite some contemporary progress in this area, accurate modeling of the unclosed terms that are predictive across regimes remains an outstanding challenge. Fox [39] proposed closures of the phase-averaged (PA) terms based largely on single-phase turbulence models without extensive validation. Capecelatro et al. [54] extended these models to account for near-wall effects in particle-laden channel flows. Agreement with the turbulence statistics obtained from simulation data was found to be satisfactory at first order (e.g. PA velocities) but less so at second order (e.g. PA turbulent kinetic energy). More recently, Innocenti et al. [55] drew upon a probability-density-function approach, along with extensions from single-phase turbulence modeling (particularly in the fluid phase), showing satisfactory agreement for statistics up to second order. However, the model was again restricted to relatively dilute flows. Due to the complex physics, wide range of regimes present and the large parameter space associated with turbulent, disperse two-phase flows, a reliable modeling approach valid across regimes remains elusive.

Thus, the goal of this dissertation is to develop a modeling framework capable of taking highly resolved data as input and produce multiphase RANS models that are accurate across regimes,interpretable and easily shared across users and solvers. A secondary goal is to employ this methodology to propose improved models for turbulent, gas-solid flows.

1.3 Objectives of this work

The primary goal of this work is to close the gap between physical observations and the predictions made by existing models in disperse two-phase systems, such as

circulating fluidized bed reactors. Due to the large span of length- and time-scales present in industrial systems, the RANS equations continue to be the workhorse of industry. However, currently models which accurately account for the effect of heterogeneity, like clustering, do not exist. For this reason, broad assumptions are made in the RANS equations, such as uniformity in the disperse phase and perfect mixing. These assumptions, which neglect the rich and complex physics present, are shown to have substantial errors such as underpredicting the time [56] or distance [57] to thermal equilibrium or over predicting the conversion rate of feedstock [6]. These errors directly lead to sub-optimal designs of physical systems and operating conditions.

Thus, the goal of this work is the close the gap between physical observation and reduced order models. This is done by formulating a modeling methodology that takes highly resolved data as input and generates closure that is accurate across scales and parameters. Because experiments are expensive and difficult, particularly in dense regimes, this work leverages high fidelity simulations that capture mesoscale behavior, such as clustering, to serve as input to the modeling methodology. Then, the highly resolved data and modeling methodology are married to produce reduced order models that take gas-particle physics into account and improve the prediction of key quantities in the multiphase RANS equations.

The specific objectives of this dissertation include:

1. Illustrate the need for reduced order models in turbulent, gas-solid flows by quantifying the effect of heterogeneity on the pyrolytic conversion of biomass to biofuel.

2. Develop a modeling methodology for distilling highly resolved data into closures that are accurate, interpretable, transportable and easy to use.

3. Demonstrate the viability of the methodology on several canonical flows, begin-

ning with single-phase turbulence.

4. Extend the modeling framework to multiphase flows of increasing complexity and dimensionality.

1.4 Thesis overview

This dissertation is divided into seven chapters. Chapters II-VI are reorganized manuscripts that are either already in print or have been submitted for publication in a journal.

Chapter II [6] presents an example that motivates the importance of developing improved models for gas-solid flows.

Chapter III [58] details the development of a modeling methodology termed 'Sparse Regression with Embedded Invariance.' This method leverages highly resolved simulations and distills this information into improved models with the important properties discussed previously: interpretability, ease of dissemination and ease of integration into existing solvers. In this chapter, the proposed methodology is applied to several canonical single phase flows (homogeneous free shear turbulence, turbulence over periodic hills and turbulence over a backward facing step) as an initial test of concept.

Chapter IV [59] extends the methodology presented in Chapter III to gas-solid flows by examining one of the simplest flow configurations in which strong coupling between the phases generates and sustains turbulence in the fluid phase. Specifically, each of the terms appearing in the transport of the multiphase Reynolds stress tensors is modeled and a full set of RANS equations is proposed. These closures are then demonstrated to have excellent performance on a temporally evolving set of test flows.

Next, Chapter V [57] extends the scope and dimensionality of modeling multiphase flows by considering thermally evolving gas-solid flow. Particular attention is paid

to predicting the thermal entrance length, the distance after which the phases are in a statistical thermal equilibrium with each other. In this chapter, the effect of heterogeneity on the thermal entrance length is quantified and a model that takes these effects into account is proposed.

In Chapter VI [60], an improvement to the sparse regression modeling methodology is developed by integrating gene expression programming into the modeling workflow. This blended approach retains the benefits of sparse regression while also automating the modeling process in situations where physical arguments cannot be used to guide trial function selection.

Finally, Chapter VII summarizes the major findings and contributions of the work presented and lays out a glide path for future studies.

CHAPTER II

A Motivating Example: Biomass Pyrolysis in Fully Developed Riser Flow

This chapter presents a numerical study of biomass pyrolysis in turbulent riser flow. Eulerian–Lagrangian simulations of unbounded sedimenting gas-solid flows are performed to isolate the effects of particle clustering on the production of syngas and tar. This configuration provides a framework to resolve the relevant length- and time-scales associated with thermal, chemical and multiphase processes taking place in the fully-developed region of a circulating fluidized bed riser. A four-step kinetic scheme is employed to model the devolatilization of biomass particles and secondary cracking of tar. Two-way coupling between the phases leads to clusters of sand particles that generate and sustain gas-phase turbulence and transport biomass particles. Neglecting the heterogeneity caused by clusters was found to lead to a maximum over-prediction of syngas yield of 33%. Further, it was found that two-dimensional simulations over-predict the level of clustering, resulting in an under-prediction of syngas and tar yields.

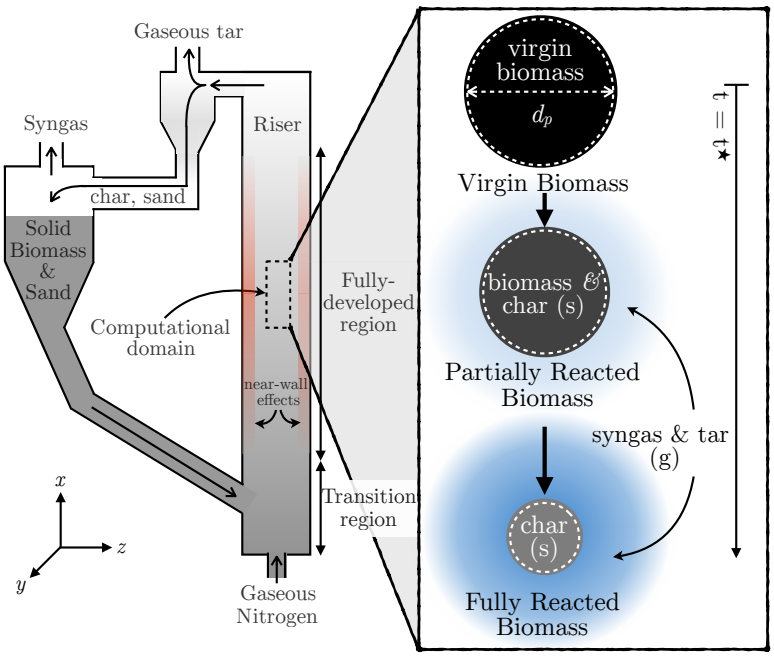


Figure 2.1: Schematic of a representative biomass pyrolysis reactor. The dashed box represents the simulation domain considered in the present study, which lies in the fully-developed region of the riser sufficiently far from wall effects.

2.1 Introduction

Biofuels generated from the pyrolytic upgrading of biomass are becoming a formidable alternative to fossil fuels, with the important benefits of sustainability, carbon neutrality and economic feasibility [61, 62, 63, 64]. The catalytic upgrading of biomass solids (e.g. cellulose-based materials often originating from agricultural waste) to useful products such as syngas, typically occurs in the riser of a circulating fluidized bed reactor (see Figure 2.1). While such technologies show great promise in a laboratory setting, scale-up to commercial application remains a key hurdle [65]. A recurring challenge is the need to ensure optimal contact between the reacting flow and fluidized particles (e.g. catalysts and biomass). While there has been significant progress characterizing hydrodynamic interactions in particle-laden flows [66], much less is known about interphase heat and mass transfer.

In a circulating fluidized bed reactor, biomass is injected in the riser, and the

feedstock and catalyst flow upward. Two-way coupling between the turbulent flow and fluidized solids leads to the spontaneous generation of densely packed particles, referred to as clusters [41, 67, 68, 69]. Such large-scale structures have been observed to 'demix' the underlying flow [4], resulting in reduced contact between the phases with a direct consequence on heat and mass transfer [4, 70, 71, 56]. Due to the vast separation in length- and time-scales–from the scale of individual particles to reactor-scale processes–in addition to reduced optical access caused by the particles, experimental studies of biomass pyrolysis in riser flows have proven to be an arduous task.

In recent years, particle-resolved direct numerical simulation (PR-DNS) has emerged as a promising technique to understand and model reacting particle-laden flows [72, 73, 74, 75, 5, 76, 77, 78]. In PR-DNS, the resolution of the computational grid is significantly smaller than the particle diameter such that all of the relevant thermo-chemical/hydrodynamic scales are properly resolved. While studies like these provide critical insight on detailed processes occurring at the particle scale, the high computational cost of these methods prevent state-of-the-art PR-DNS from capturing emerging *mesoscale* structures (e.g. clusters) that may span hundred to thousands of particles. To this end, alternative approaches are typically employed. These are commonly distinguished between Eulerian–Eulerian (EE) methods that treat both the particle and gas phase on a common Eulerian grid [79], and Eulerian–Lagrangian (EL) methods [80, 81], which solves the fluid on a computational grid and tracks individual particles.

The application of EE methods to biomass pyrolysis began appearing in the literature only within the last decade. Xiong et al. [82] used a multi-fluid model combined with a four-step kinetic scheme to simulate bagasse in a laboratory-scale two-dimensional domain. The simulation was found to slightly under-predict tar formation and slightly over-predict char and syngas. Xue et al. [83] performed simulations of

a similar configuration. Shrinkage of the biomass was accounted for via an apparent density model, which provided reasonable predictions for laboratory-scale biomass conversion. Boateng and Mtui [84] used an EE method combined with a one-step kinetic scheme to simulate switchgrass, corn cobs and soy bean biomass sources in a three-dimensional laboratory-scale reactor. They found reasonable success when comparing with experimental measurements, however they noted that a generalized model should be developed due to the large degree of variability in biomass composition in physical systems.

EL strategies provide an alternative framework that typically rely on simpler closure modeling compared to EE, where individual particle trajectories are solved using Newton's laws of motion, and models are required for interphase exchange and particle collisions. Because particles are tracked individually, particle trajectory crossing is naturally captured [85], which plays a large role in turbulent particle-laden flows and was recently found to control the onset of clustering [86]. Only very recently EL methods have been used to simulate biomass pyrolysis [87, 88, 89, 90, 91, 92, 93, 71, 94]. These have mostly focused on full-scale reactors, requiring special treatment for the inlet, outlet, and reactor walls. Meanwhile, simple, canonical flow configurations that isolate multiphase flow dynamics on biomass conversion have yet to be proposed. The focus of the present study is to introduce a framework to model the fully-developed region of a riser, and capture the detailed thermochemical processes in the presence of clusters.

In the following sections, an EL approach to modeling biomass pyrolysis in three-dimensional, unbounded, fully-developed riser flows will be presented. Of particular interest is the role of heterogeneity on pyrolytic conversion of biomass particles. We draw comparisons between a computation with a zero-dimensional, homogeneous model (i.e. non-clustered) and an equivalent system in which clusters form and reach a statistically stationary state. We make these comparisons in both two- and three-

dimensions and comment on the impact of dimensionality on pyrolytic yields.

2.2 Methodology

2.2.1 System description

A representative schematic of a biomass reactor is shown in Figure 2.1 and the computational domain is shown in greater detail in Figure 2.4. In this system, hot sand is fluidized by heated nitrogen and cold biomass particles are injected into the riser. As the biomass particles are heated, they reach an activated state and subsequently devolatilize. This process results in the production of gaseous tar and syngas and the breakdown of biomass to char.

The simulation configuration considered in the present study is designed to provide a model flow that captures key phenomenology in the reactor. We focus on what would be considered the fully-developed region of the riser. Assuming the flow is sufficiently far from the entrance, the two phases would evolve spatially upward without retaining memory of any entrance effects. If we further assume the simulation domain is far from the walls of the riser, and given that the flow equations are frame invariant, we can consider a temporally-developing frame of reference in an unbounded system. Focusing on the fully-developed region of the riser flow enables us to resolve a broader range of length- and time-scales associated with biomass pyrolysis than could otherwise be achieved. Further, the present configuration isolates the role of multiphase dynamics on production yield.

All relevant simulation parameters are summarized in Table 2.1. To ensure the hydrodynamics are independent of the domain size, the simulation configuration corresponds to Case 4 of [98]. In this case, the ratio of grid spacing to the diameter of the sand particle is $\Delta x/d_p = 1.5$. We note that the biomass is initially larger than the grid spacing, which is enabled via a two-step projection method described in [81].

Physical properties					
			biomass	*sand*	*char*
d_p	[µm]	Particle diameter	500	200	
ρ_p	[kg/m^3]	Particle density[95]	400	2649	2333
$C_{p,p}$	[J/kg K]	Particle heat capacity[96]	2300	800	1100
T_p^0	[K]	Initial particle temperature	300	790	
κ_p	[J/m s K]	Particle thermal conductivity[96]	0.3	0.27	1
			N_2	*tar*	*syngas*
ρ_g	[kg/m^3]	Fluid density	EOS[1]	EOS[1]	EOS[1]
$C_{p,g}$	[J/kg K]	Fluid heat capacity[96]	1121	2500	1100
T_g^0	[K]	Initial fluid temperature	790	–	–
κ_g	[J/m s K]	Fluid thermal conductivity[96]	5.63×10^{-2}	2.57×10^{-2}	
ν_g	[m^2/s]	Fluid kinematic viscosity	[97]	$3\times10^{-5}\rho_g$ [96]	
Dimensional parameters					
			biomass	*sand*	
τ_p	[s]	Particle response time	0.15	0.16	
$\mathcal{L}$	[m]	Cluster length	0.004	0.005	
$\mathcal{V}$	[m/s]	Terminal velocity	0.028	0.0297	
$\boldsymbol{g}$	[m/s^2]	Gravity	$(-0.18, 0, 0)$		
Non-dimensional parameters					
			biomass	*sand*	
$\langle\varepsilon_p\rangle$		Mean particle volume fraction	0.005	0.01	
N_p		Number of particles	17 302	540 710	
Φ_m		Mass loading	4.74	62.8	
Re		Reynolds number	1.46	1.74	
Fr		Froude number	8.71	24.4	
Ar		Archimedes number	3.01	1.28	

Table 2.1: Summary of relevant simulation parameters. [1] Equation of state for an ideal gas (Eq. 2.19). Note all biomass quantities are given at the unreacted state (i.e. at $t = t^\star$).

The domain is triply periodic with $512 \times 128 \times 128$ uniformly spaced grid points in a domain of length $31.4\mathcal{L} \times 7.8\mathcal{L} \times 7.8\mathcal{L}$, where $\mathcal{L} = \tau_p^2 g$ is an *a priori* measure of cluster size [67, 99, 100, 41]. Here, $\tau_p = \rho_p d_p^2/(18\mu_g)$ is the particle response time, with ρ_p the particle density, μ_g the gas-phase viscosity, and g is gravity. Here, the magnitude of gravity has been chosen as 0.18 in order to set the relevant non-dimensional numbers such that the hydrodynamics are properly resolved (see,[98]). Unless otherwise noted, the subscript 'p' denotes a sand particle throughout the remainder of the paper.

Both sand and biomass particles are initially randomly distributed throughout the domain with a prescribed volume fraction of .01 and .005, respectively, representative of typical riser flow. During the initial transient, the two phases are taken to be inert and isothermal. Particles fall under gravity and entrain the gas phase. Two-way coupling between the phases results in the spontaneous generation of particle clusters. These clusters further entrain the gas phase downward while the gas flows upward in regions void of particles, a process known as 'jet bypassing.' Eventually, the average cluster size reaches a steady value and wakes shed by clusters sustain velocity fluctuations in the fluid, referred to as fully-developed cluster-induced turbulence (CIT) [40, 101, 41]. At this point, pyrolytic reactions are activated, heat and mass transfer are permitted between the phases and transient data is collected until a second steady state is reached at which time all of the biomass has been expended.

CIT is governed by several important non-dimensional parameters. The Reynolds number dictates the level of turbulence, defined as

$$\mathrm{Re} = \frac{\rho_g \mathcal{V} \mathcal{L}}{\mu_g} \tag{2.1}$$

where $\mathcal{V} = \tau_p g$ denotes the terminal velocity of an isolated particle in a corresponding flow, and ρ_g is the gas-phase density. Additionally, the Archimedes number, $\mathrm{Ar} = (\rho_p/\rho_g - 1) d_p^3 g/\nu_g^2$, where ν_g is the fluid kinematic viscosity, is a common measure

used to characterize inertial forces in sedimenting gas-solid flows (see e.g. [102, 100]). The mean mass loading informs the importance of two-way coupling, given by

$$\Phi_m = \frac{\rho_p \langle \varepsilon_p \rangle}{\rho_g \langle \varepsilon_g \rangle}, \tag{2.2}$$

where ε_p is the particle volume fraction, $\varepsilon_g = 1 - \varepsilon_p$ is the gas-phase volume fraction, and angled brackets denote a volume average. In typical riser flows, $\Phi_m \gg 1$, and thus particles are expected to largely influence the underlying turbulence. Finally, the Froude number, $\mathrm{Fr} = \tau_p^2 g / d_p$, characterizes the competition between gravitational and inertial forces. These values are summarized in Table 2.1.

2.2.2 Chemical kinetics

During catalytic upgrading, biomass particles are decomposed into char, syngas and tar vapor. The tar vapor undergoes secondary cracking to form additional syngas. In the current study, this process is modeled by four irreversible, first-order reactions commonly used in the literature [83, 82]. A summary of the kinetic scheme is shown in Figure 2.2. Here, solid biomass is composed of three components: cellulose, hemicellulose and lignin. The biomass under consideration in this study is bagasse, the woody pulp bi-product resultant of the commercial processing of sugarcane. The species composition of bagasse is given as

$$\mathrm{Biomass} = \alpha \mathrm{Cellulose} + \beta \mathrm{Hemicellulose} + \gamma \mathrm{Lignin}, \tag{2.3}$$

where the initial composition is given by $(\alpha, \beta, \gamma) = (0.36, 0.47, 0.17)$ [83].

Upon heating, virgin biomass is activated and then devolatized into char, tar and syngas. Char remains within the particle while tar and syngas are transferred to the gas phase. The mass rates of change for the species involved in the activation and

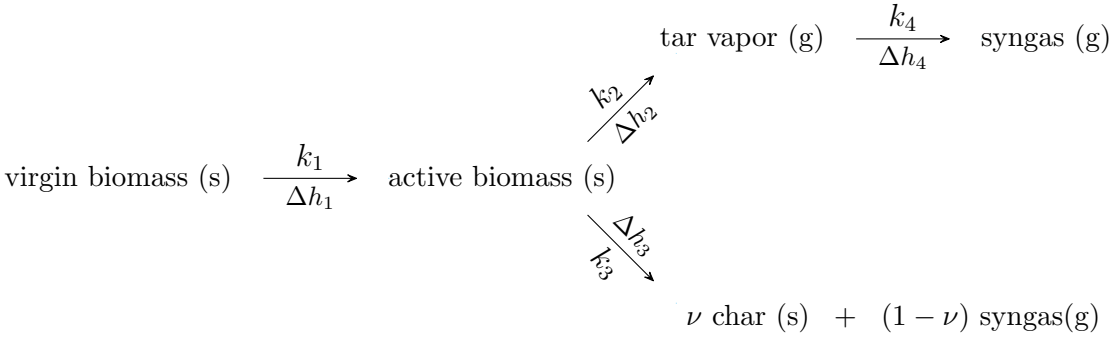


Figure 2.2: Kinetic scheme used in the present work. Solid biomass consists of cellulose, hemicellulose and lignin. The phase of each constituent is noted in parentheses. The heats of formation for each reaction are given by: $\Delta h_1 = 0$ [7], $\Delta h_2 = 255$ kJ/kg [8], $\Delta h_3 = -20$ kJ/kg [8], and $\Delta h_4 = -42$ kJ/kg [9].

devolatization of biomass particles are given by

$$\frac{dm_i^v}{dt} = -k_{1_i} m_i^v \tag{2.4}$$

$$\frac{dm_i^a}{dt} = k_{1_i} m_i^v - (k_{2_i} + k_{3_i}) m_i^a \tag{2.5}$$

$$\frac{dm_{\text{char}}}{dt} = \sum_i \nu_i k_{3_i} m_i^a \tag{2.6}$$

$$\frac{dm_{\text{tar}}}{dt} = \sum_i k_{2_i} m_i^a \tag{2.7}$$

$$\frac{dm_{\text{syn}}}{dt} = \sum_i (1-\nu_i) k_{3_i} m_i^a \tag{2.8}$$

where $i = c, h, l$ corresponds to the biomass species cellulose, hemicellulose and lignin and the superscripts a and v refer to activated and virgin biomass, respectively. Secondary cracking, as it is a homogeneous reaction in the gas phase, is solved exclusively on the Eulerian mesh. Since the biomass particles are composed at any point in time of virgin biomass, active biomass and char, the previous equations allow for an expression for the total rate of mass change for a biomass particle to be written as

$$\frac{dm_{\text{bio}}}{dt} = \sum_i \frac{dm_i^v}{dt} + \sum_i \frac{dm_i^a}{dt} + \frac{dm_{\text{char}}}{dt}. \tag{2.9}$$

Constituent	Reaction	ν (k_3)	A (s^{-1})	E (MJ/kmol)
Cellulose [103]	k_{1c}	-	2.8×10^{19}	242.4
	k_{2c}	-	3.28×10^{14}	196.5
	k_{3c}	(ν_c) 0.35	1.3×10^{10}	150.5
Hemicellulose [104]	k_{1h}	-	2.1×10^{16}	186.7
	k_{2h}	-	8.75×10^{15}	202.4
	k_{3h}	(ν_h) 0.60	2.6×10^{11}	145.7
Lignin[104]	k_{1l}	-	9.6×10^{8}	107.6
	k_{2l}	-	1.5×10^{9}	143.8
	k_{3l}	(ν_l) 0.75	7.7×10^{6}	111.4
Tar [105]	k_4	-	4.25×10^{6}	108.0

Table 2.2: Rate coefficients used in the Arrhenius reaction (2.10).

Reaction rates for the kinetics are dictated by the Arrhenius equation

$$k_i = A_i \exp[-E_i/RT_p] \tag{2.10}$$

where A_i is the pre-exponential factor, E_i is the activation energy, R is the gas constant, and T_p is the particle temperature. These quantities are summarized in Table 2.2. Chemical kinetics are treated on the scale of the particle, and as such the mass of each species contained within each particle is tracked, but the transport within the particle itself is not considered.

2.2.3 Gas-phase description

The gaseous phase initially consists of pure nitrogen and evolves into a mixture of nitrogen, gaseous tar and syngas. To enforce conservation of mass, momentum and energy in the presence of solid particles, we consider the volume-filtered Navier–Stokes equations [106, 81]. Conservation of mass is given by

$$\frac{\partial}{\partial t}(\varepsilon_g \rho_g) + \nabla \cdot (\varepsilon_g \rho_g \boldsymbol{u}_g) = \sum_{l=1}^{2} \dot{\mathcal{M}}_l \tag{2.11}$$

where $\boldsymbol{u}_g = [u_g, v_g, w_g]^{\mathrm{T}}$ is the fluid velocity and $\mathcal{M}_l$ is the interphase mass source term, with $l = (\mathrm{tar}, \mathrm{syn})$, referring to the reactive species, tar and syngas, respectively. Conservation of momentum is given by

$$\frac{\partial}{\partial t}(\varepsilon_g \rho_g \boldsymbol{u}_g) + \nabla \cdot (\varepsilon_g \rho_g \boldsymbol{u}_g \boldsymbol{u}_g) = \varepsilon_g \nabla \cdot \boldsymbol{\tau}_g + \varepsilon_g \rho_g \boldsymbol{g} + \mathcal{F} + \boldsymbol{F}_{\mathrm{mfr}} \tag{2.12}$$

where $\mathcal{F}$ accounts for two-way coupling between the gas and solid phases, which will be made explicit in Section 2.2.5. $\boldsymbol{F}_{\mathrm{mfr}}$ is a source term used to ensure that the system maintains a constant mass flow rate in order to achieve a statistically stationary state. The viscous stress tensor is given by

$$\boldsymbol{\tau}_g = -p_g \mathbb{I} + \mu_g^* \left[\nabla \boldsymbol{u}_g + \nabla \boldsymbol{u}_g^{\mathrm{T}} - \frac{2}{3} (\nabla \cdot \boldsymbol{u}_g) \mathbb{I} \right], \tag{2.13}$$

where p_g is the gas-phase pressure, $\mathbb{I}$ is the identity matrix and μ_g^* is an effective viscosity accounting for enhanced dissipation at the particle scale, given by [107]

$$\mu_g^* = \mu_g \varepsilon_g^{-2.8}. \tag{2.14}$$

The species transport equations are given by

$$\frac{\partial}{\partial t}(\varepsilon_g \rho_g Y_l) + \nabla \cdot (\varepsilon_g \rho_g Y_l \boldsymbol{u}_g) = \varepsilon_g \nabla \cdot (\rho_g \mathcal{D} \nabla Y_l) + \dot{\omega}_l + \dot{\mathcal{M}}_l \tag{2.15}$$

where Y_l is the mass fraction of each species. To ensure global conservation, the mass fraction of nitrogen is defined as $Y_{N_2} = 1 - (Y_{\mathrm{syn}} + Y_{\mathrm{tar}})$. Here, $\mathcal{D}$ is the mass diffusivity, which is determined by maintaining a constant Schmidt number $\mathrm{Sc} = \nu_g / \mathcal{D} = 0.7$. Since the only gaseous homogeneous reaction under consideration is that of tar to syngas conversion, $\dot{\omega}_l$ is given by

$$\dot{\omega}_{\mathrm{syn}} = -\dot{\omega}_{\mathrm{tar}} = \varepsilon_g \rho_g k_4 Y_{\mathrm{tar}} \tag{2.16}$$

and $\dot{\omega}_{N_2} = 0$. The final term in Eq. (2.15) accounts for mass transfer from solid biomass to gas species, which will be defined in Section 2.2.5.

Energy conservation is expressed as

$$\frac{\partial}{\partial t}(\varepsilon_g \rho_g C_{p,g} T_g) + \nabla \cdot (\varepsilon_g \rho_g C_{p,g} T_g \boldsymbol{u}_g) = \varepsilon_g \nabla \cdot (\kappa_g^* \nabla T_g) + \mathcal{Q} + \dot{\omega}_T + \dot{Q} \tag{2.17}$$

where $\mathcal{Q}$ is the interphase heat exchange term that will be defined later, $\dot{\omega}_T = \Delta h_4 \dot{\omega}_{syn}$ accounts for the heat of formation of the secondary cracking of syngas, and κ_g^* is an effective thermal conductivity that accounts for enhanced dissipation of heat at the particle scale, defined as [108]

$$\kappa_g^* = \frac{1 - \sqrt{1 - \varepsilon_g}}{\varepsilon_g} \kappa_g. \tag{2.18}$$

In this work, the thermal conductivity, κ_g, is determined by maintaining a constant Prandtl number $\mathrm{Pr} = \mu_g C_{p,g} / \kappa_g = 0.7$.

It should be noted that the initial temperature difference between the gas phase and biomass particles will result in transient cooling that would directly affect the rate of reactions. In real systems, biomass particles exchange heat with fresh gas as they are transported spatially in the riser. To mimic this, a uniform heat source $\dot{Q}$ is added to the energy equation to maintain a constant mean temperature $\langle T_g \rangle = 790$ K. As a result, the gas phase is expected to cool near the biomass particles and heat up away from them. Further details on enforcing thermally fully-developed flows can be found in [56].

The gas-phase equations are solved in NGA [85], a fully conservative, low-Mach number solver. A pressure Poisson equation is solved to enforce continuity (2.11) via fast Fourier transforms in all three periodic directions. As a consequence, the

gas-phase density is determined from the equation of state

$$\rho_g = \frac{P_g \overline{W}}{R_u T_g}, \tag{2.19}$$

where P_g is the thermodynamic pressure, $R_u = 8.314$ J/mol·K is the universal gas constant and $\overline{W}$ is the mixture molecular weight defined as

$$\frac{1}{\overline{W}} = \sum_{k=1}^{N} \frac{Y_k}{W_k}. \tag{2.20}$$

The Navier-Stokes equations are solved on a staggered grid with second order spatial accuracy and time is advanced with second order accuracy using the semi-implicit Crank-Nicolson scheme of Pierce [109]. Advection of temperature and mass fraction is solved using the bounded quadratic-upwinded biased interpolative connective (BQUICK) scheme [110].

2.2.4 Particle-phase description

All particles are assumed to be spherical and rigid with an initial diameter of 500 μm (for biomass) and 200 μm (for sand) as shown in Table 2.1. Particles are tracked individually in a Lagrangian manner according to Newton's second law of motion, given by

$$\frac{d\boldsymbol{x}_p}{dt} = \boldsymbol{u}_p \tag{2.21}$$

and

$$\frac{d}{dt}\left(m_p \boldsymbol{u}_p\right) = V_p \nabla \cdot \boldsymbol{\tau}_g[\boldsymbol{x}_p] + \boldsymbol{f}_{\text{drag}} + \frac{dm_p}{dt}\boldsymbol{u}_p + \mathbb{C} + m_p \boldsymbol{g}, \tag{2.22}$$

where $\boldsymbol{x}_p$ and $\boldsymbol{u}_p$ are the position and velocity of particle p, respectively, m_p is the particle mass, and V_p is its volume. Square brackets denote a fluid quantity interpolated

to the center position of particle p. The drag force is defined as

$$\boldsymbol{f}_{\text{drag}} = m_p \frac{\varepsilon_g}{\tau_p}(\boldsymbol{u}_g[\boldsymbol{x}_p] - \boldsymbol{u}_p)F(\varepsilon_g, \text{Re}_p) \tag{2.23}$$

where Re_p is the particle Reynolds number, given by

$$\text{Re}_p = \frac{\varepsilon_g \rho_g \left|\boldsymbol{u}_g[\boldsymbol{x}_p] - \boldsymbol{u}_p\right| d_p}{\mu_g} \tag{2.24}$$

and $F(\varepsilon_g, \text{Re}_p)$ is a non-dimensional correction factor to account for volume fraction and Reynolds number effects [111]. The force due to inter-particle collisions, $\mathbb{C}$, is modeled using a modified soft-sphere approach originally proposed by Cundall and Strack [112]. Particles are treated as inelastic and frictional with a coefficient of restitution of 0.85 and coefficient of friction of 0.1. The interested reader is referred to Capecelatro and Desjardins [81] for additional details.

The evolution of solid-phase temperature is given by

$$m_p C_{p,p} \frac{dT_p}{dt} = V_p \nabla \cdot \left(\kappa_g^* \nabla T_g[\boldsymbol{x}_p]\right) + q_{\text{inter}}, \tag{2.25}$$

where $C_{p,p}$ is the particle-phase heat capacity. The first term on the right-hand side of (2.25) accounts for 'resolved' heat fluxes at the particle location. The last term accounts for 'unresolved' heat fluxes due to particle-scale convection and devolatilization, defined as

$$q_{\text{inter}} = q_{\text{conv}} + q_{\text{devol}}. \tag{2.26}$$

In the present study, inert sand particles only exchange heat via convection and the biomass particles have contributions from both terms. The interphase heat exchange due to convection is expressed as

$$\frac{q_{\text{conv}}}{V_p} = \frac{6\kappa_g \text{Nu}_p}{(d_p)^2}\left(\widetilde{T}_g - T_p\right) \tag{2.27}$$

where the Nusselt number, Nu, is computed using Gunn's correlation [113]. The heat exchange term due to devolatilization is given by

$$q_{\text{devol}} = \left[\Delta h_2 \frac{dm_{\text{tar}}}{dt} + \Delta h_3 \left(\frac{dm_{\text{char}}}{dt} + \frac{dm_{\text{syn}}}{dt}\right)\right] \tag{2.28}$$

noting that the heat of formation for the secondary cracking of tar is treated in the gas-phase equations.

It should also be noted that the present formulation assumes a uniform temperature distribution throughout the particle. This assumption can be justified when the Biot number is small, defined as Bi= $hd_p/6\kappa_p$, where κ_p is the thermal conductivity of the particle and h =Nuκ_g/d_p is the heat transfer coefficient. Using the values in Table 2.1 to estimate the average Nusselt number, the Biot number is expected to remain below 0.09 for biomass and sand, which supports the uniform particle temperature assumption.

Finally, the shrinkage of biomass is explicitly accounted for as it is converted to char, tar and syngas. The mass of a biomass particle, m_{bio}, is equal to the sum of the mass of each of its constituents, i.e.

$$m_{\text{bio}} = \sum_i m_i^v + \sum_i m_i^a + m_{\text{char}}, \tag{2.29}$$

where as before, $i = c, h, l$ corresponds to cellulose, hemicellulose and lignin and the superscripts a and v refer to activated and virgin biomass, respectively. Additionally, the density of the particle is given by

$$\rho_{\text{bio}} = \left(\frac{(\sum_i Y_i^v + \sum_i Y_i^a)}{\rho_{\text{bio}}} + \frac{Y_{\text{char}}}{\rho_{\text{char}}}\right)^{-1}, \tag{2.30}$$

where ρ_{bio} and ρ_{char} are provided in Table 2.1. With this, the volume of biomass is given by $V_{\text{bio}}(t) = m_{\text{bio}}/\rho_{\text{bio}}$ and the resulting diameter of the spherical particle can

be determined according to $d_{\text{bio}}(t) = (6V_{\text{bio}}(t)/\pi)^{1/3}$.

2.2.5 Two-way coupling

The gas-phase equations introduced in Section 2.2.3 contain several interphase exchange terms that require careful attention. These terms include the particle volume fraction, ε_p, mass exchange, $\mathcal{M}_l$, for each species l, momentum exchange, $\mathcal{F}$, and heat exchange, $\mathcal{Q}$. All of these terms require projecting information from Lagrangian particles to the computational mesh. To accomplish this in an efficient and conservative manner, the two-step filtering approach of Capecelatro and Desjardins [81] is employed. First, particle data is extrapolated to the nearest grid points. Then, a second 'smoothing' operation is performed implicitly, such that the final support of the filtering operation is tied to a chosen filter size δ_f. Here, we consider a Gaussian filter kernel $\mathcal{G}$ with $\delta_f = 7$ times the mean particle diameter.

The gas-phase volume fraction is computed according to

$$\varepsilon_g = 1 - \sum_{p=1}^{N_p} \mathcal{G}\left(|\boldsymbol{x} - \boldsymbol{x}_p|\right) V_p, \tag{2.31}$$

where N_p is the total number of particles (biomass and sand).

Similarly, mass exchange is defined as

$$\dot{\mathcal{M}}_l = \sum_{p=1}^{N_p} \mathcal{G}(|\boldsymbol{x} - \boldsymbol{x}_p|) \frac{dm_l}{dt} \tag{2.32}$$

where l refers to tar and syngas. Since sand is non-reactive, it naturally does not contribute to the interphase mass transfer term.

The momentum exchange term is given by

$$\mathcal{F} = -\sum_{p=1}^{N_p} \mathcal{G}\left(|\boldsymbol{x} - \boldsymbol{x}_p|\right) \left(\boldsymbol{f}_{\text{drag}} + \frac{dm_p}{dt} \boldsymbol{u}_p \right). \tag{2.33}$$

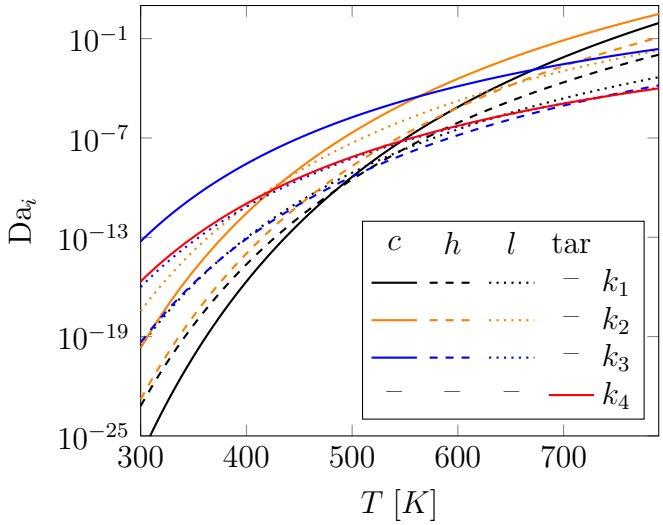


Figure 2.3: Dependence of Damköhler number on temperature. Biomass constituents, cellulose, hemicellulose and lignin, are denoted as c, h and l, respectively, in the legend.

Finally, the heat exchange term is given by

$$\mathcal{Q} = -\sum_{p=1}^{N_p} \mathcal{G}\left(|\boldsymbol{x} - \boldsymbol{x}_p|\right) q_{\text{inter}}. \tag{2.34}$$

2.2.5.1 A note on the Damköhler number

The Damköhler number is used to assess the relative importance between fluid dynamic time scales and and pyrolytic reactions, and provides an *a-priori* estimate of the rate limiting processes. Due to the simplicity of the configuration considered in the present work (i.e., unbounded flow with zero mean mass flow rate) there exists few characteristic time scales to choose from. Taking τ_{bio}, the particle response time of biomass, to serve as the characteristic fluid time scale, and the inverse of the reaction rate as the kinetic time scale, the Damköhler number can be expressed as

$$\mathrm{Da}_i = k_i \tau_{\text{bio}} \tag{2.35}$$

where i indicates each of the 10 reaction rates listed in Table 2.2. It is important to note that when the chemistry is slow (i.e. $\mathrm{Da}_i \ll 1$), particle clustering and two-way coupling between the phases will play a dominating role. For values $\mathrm{Da}_i \gg 1$, chemical kinetics will appear to take place in an approximately 'frozen' flow field.

Because the reaction rates are functions of temperature, Da_i also varies with temperature, as shown in Figure 2.3. At $t = t^\star$, the temperature of biomass is 300 K and will cool the gas phase relatively quickly in the vicinity of clusters. At this time, $\mathrm{Da}_i \ll 1$ and thus particle dynamics occur much faster than the kinetics. As the biomass particles are heated and begin to react, values of Da_i increase towards unity, resulting in a competition between kinetics and fluid dynamics. At elevated temperatures, conversion of active cellulose to tar is the rate-limiting reaction, while conversion of hemicellulose to char and syngas and the secondary cracking of tar remain slow compared to the particle dynamics.

2.3 Results

2.3.1 The fully-developed region of a riser flow

After an initial transient of approximately $100\tau_p$, the flow reaches a statistically stationary state where turbulence is generated and sustained by clusters. At this time (denoted by $t = t^\star$), heat and mass transfer are enabled between the phases. As shown in Figure 2.4, clusters form and entrain the gas phase downward, while high-speed jets flow upward in regions void of particles. The majority of biomass can be seen to accumulate within clusters of sand particles. It is interesting to note that while individual sand particles have similar inertia to biomass particles (refer to values of τ_p in Table 2.1), clusters–which are seen to span hundreds of particles–contain far greater inertia. As a result, the transport of biomass is largely dictated by cluster dynamics.

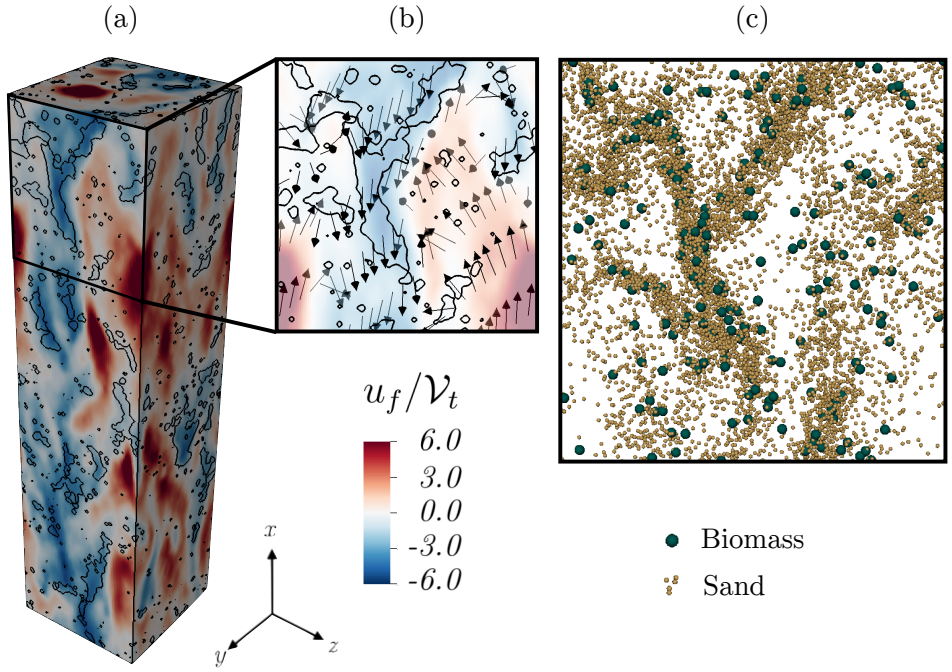


Figure 2.4: Instantaneous snapshots showing fully-developed CIT at $t = t^\star$. (a) Vertical gas-phase velocity shown in color, depicting 'jet bypassing' around clusters. (b) Fluid velocity vectors shown with black arrows. Iso-contours of $\alpha_p = 2\langle\alpha_p\rangle$ depict clusters. (c) Particle position showing sand (yellow) and biomass (green).

The level of clustering can be assessed by considering the deviation of the particle distribution from randomness [69, 98], defined by $D = (\sigma - \sigma_p)/\langle \varepsilon_p \rangle$, where σ and σ_p are the standard deviations of volume fraction for CIT and a randomly distributed particle field, respectively. At $t = t^\star$, $D = 0.62$, indicating the particle distribution is far from random. Because particles feel reduced drag within clusters, their average settling velocity exceeds their terminal velocity, $\mathcal{V}$ [114]. Once a statistically stationary state is achieved, the mean settling velocity, $\langle u_p \rangle$, is $2.0\mathcal{V}$ for biomass and $2.1\mathcal{V}$ for sand particles.

2.3.2 Thermochemical conversion

After fully-developed CIT is reached, heat and mass transfer is enabled between the phases. As discussed above, the majority of biomass particles are located within clusters of sand particles. Because biomass is initially colder than the surroundings, the gas phase cools relatively fast within clusters and heats up in regions void of particles (see Figure 5). Spatial variations in temperature then decrease as the flow evolves. It takes approximately $2\tau_p$ for biomass particles to sufficiently heat up, at which time tar and syngas are produced. The maximum mass fraction of tar is reached after approximately $5\tau_p$ and is subsequently depleted as it is cracked into syngas.

Figure 2.6 shows the evolution of Favre averaged temperature, syngas mass fraction and tar mass fraction. Here, Favre average quantities are defined as $\widetilde{(\cdot)} = \langle \varepsilon_g \rho_g (\cdot) \rangle / \langle \varepsilon_g \rho_g \rangle$. To assess the role of heterogeneity caused by clusters on thermochemical conversion, comparisons are made against a zero-dimensional system that models a homogeneous flow under identical conditions. In this system, all spatial variations are neglected and variables are solved for as a function of time only. A single, stationary biomass particle is introduced at a starting temperature of 300 K into a hot volume of nitrogen such that the volume fraction $\langle \varepsilon_p \rangle$ is consistent with the three-dimensional simulation. As was done in the three-dimensional case, the mean

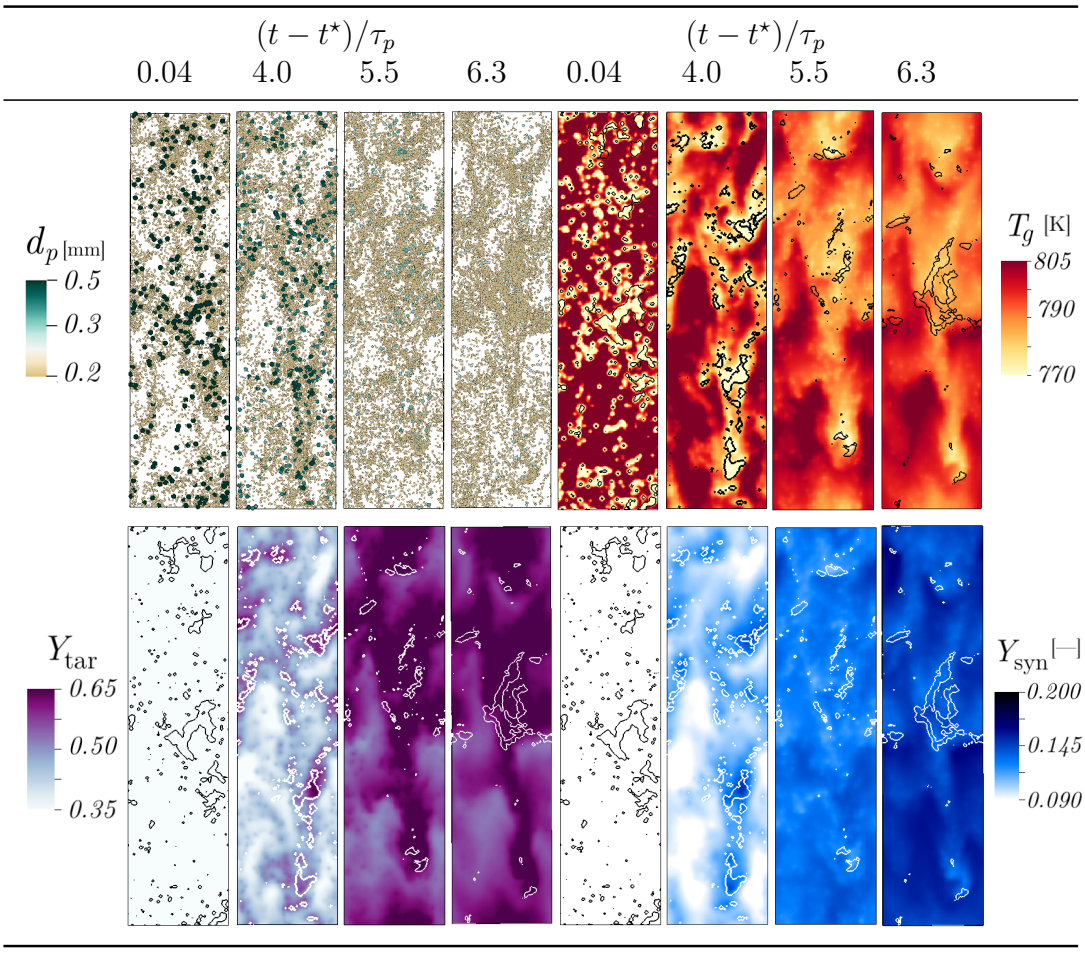


Figure 2.5: Instantaneous snapshots showing two-dimensional slices of the three-dimensional domain. Top-left: Particle position colored by diameter (sand shown in yellow, biomass shown in green). Particles are shown at 2x scale. Top-right: Gas-phase temperature. Bottom-left: Tar mass fraction. Bottom-right: Syngas mass fraction. Once the flow reaches a statistically stationary state (at $t = t^\star$), heat transfer and pyrolysis reactions are initiated. Since biomass particles are initially cooler (300K) than the sand and surrounding nitrogen (790K), the fluid surrounding the biomass is initially cooled. After approximately $2\tau_p$, the biomass particles have been sufficiently heated and tar and syngas are produced. Syngas is produced directly from the devolatization of biomass particles, in addition to the secondary cracking of tar. Thus the maximum mass fraction of tar is reached in the third column and is subsequently depleted as it is cracked into syngas. Contour lines of $\alpha_p = 2\langle\alpha_p\rangle$ depict clusters. A movie is available in the online supplementary material.

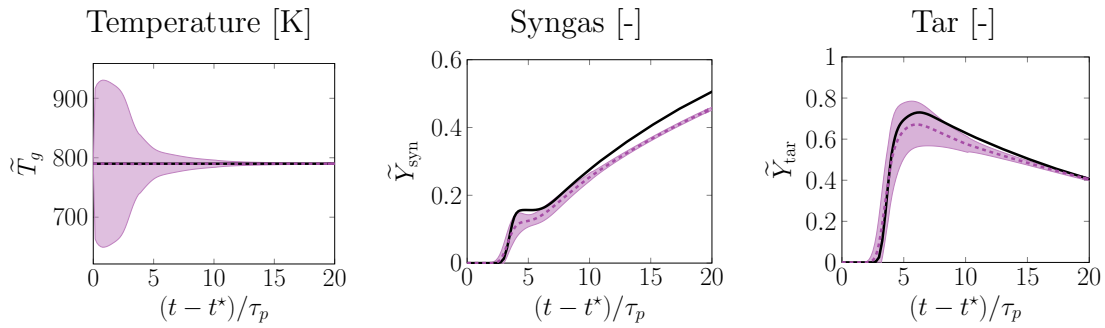


Figure 2.6: Evolution of the gas-phase temperature, syngas, and tar. Mean values for three-dimensional CIT (dashed purple lines) and a homogeneous (zero-dimensional) system under identical conditions (solid black lines). The shaded regions correspond to ± 3 times the standard deviation in CIT.

gas-phase temperature is forced to remain constant at 790 K.

Three key observations can be made: (1) spatial variations present in CIT induce variations in gas-phase temperature (which are large at early times due to the discrepancy in initial temperature between biomass particles and their surroundings). (2) Variations also exist in the mass fractions of tar and syngas. At early times ($t \lessapprox 4\tau_p$), syngas and tar are predominantly formed in clusters where the majority of biomass particles exist, despite the lower temperature in these regions. The species are then transported away from clusters, where secondary reactions occur. (3) Because the zero-dimensional model does not account for heterogeneity in spatial distribution, it over-predicts both tar and syngas formation. The typical residence time for cellulosic biomass in a circulating fluidized bed reactor is approximately 1 s, corresponding to $6.25\tau_p$. Figure 2.6 shows that the zero-dimensional model over-predicts both tar and syngas production when simulation time reaches standard residence time by 8.9% and 13.0%, respectively. It is notable that for residence times less than 1 s, the zero-dimensional model achieves a maximum over-prediction of 32.9% for syngas yield at $(t-t^\star)/\tau_p = 4.02$ (0.64 s). Maximum over-prediction in the yield of tar is 9.0% occurring at $(t-t^\star)/\tau_p = 4.63$ (0.74 s). This suggests that heterogeneity plays an especially important role in systems with short residence times, for instance flash pyrolysis.

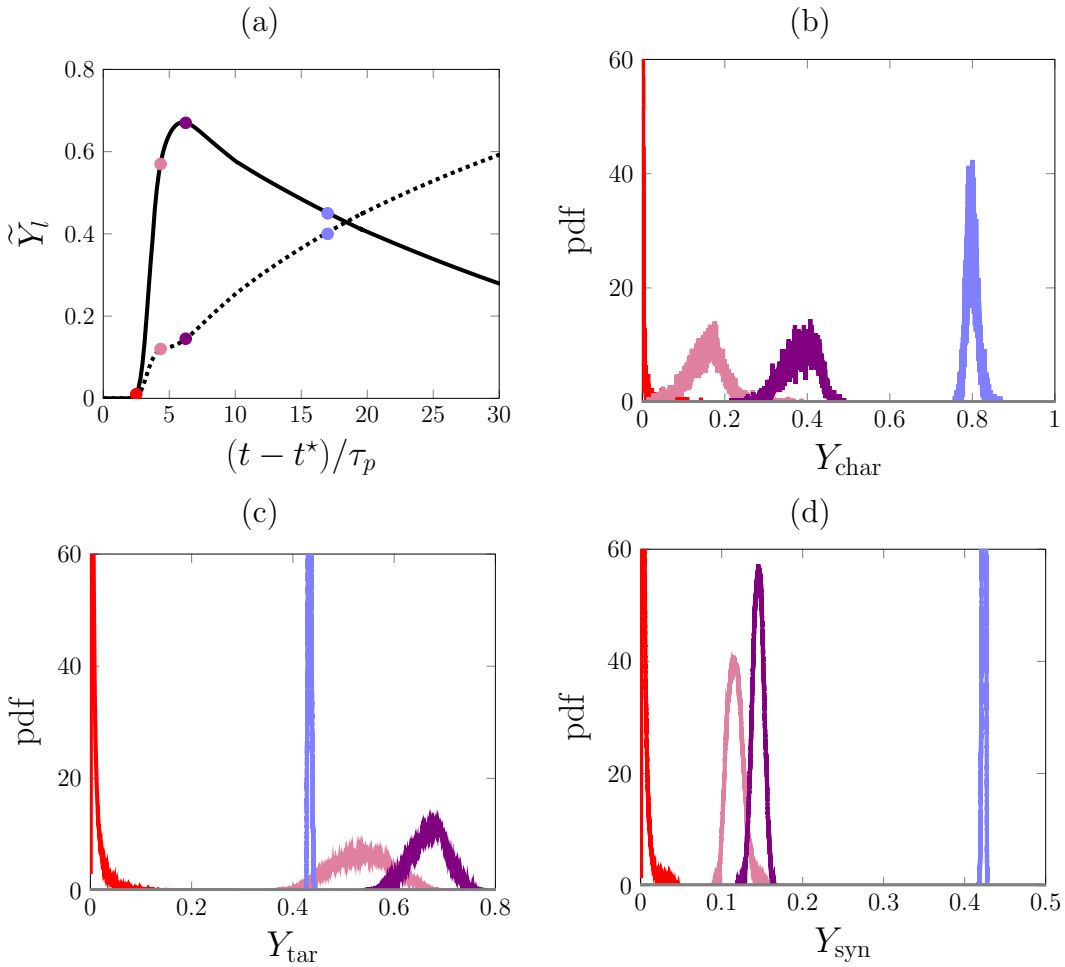


Figure 2.7: Probability distribution functions of (b) char, (c) tar, and (d) syngas. The pdfs are colored based on simulation time, as denoted in (a), showing the mean mass fractions of syngas (dotted line) and tar (solid line).

Probability distribution functions (pdfs) at four snapshots in time are provided in Figure 2.7. Char content within the biomass increases with time and its variance decreases until it eventually converges to fully reacted biomass (i.e. $\widetilde{Y}_{\text{char}} = 1$) after $20\tau_p$. Tar initially increases until it reaches a critical value, at which point there is equal competition between tar production from biomass particles and secondary cracking of tar into syngas. Then, the mean and variance of tar concentration significantly decrease. After approximately $15\tau_p$, tar concentration is uniform throughout the domain. Because there is no consumption of syngas, it increases monotonically until all the biomass and tar have been expended. In the three-dimensional CIT configuration considered here, it takes approximately $25\tau_p$ for the majority of biomass to be expended.

2.3.3 The implications of dimensionality

Industrial operations often rely on simplifications in order to make computations tractable. In this section, we briefly compare the yields corresponding to a zero-dimensional (homogeneous) configuration with two-dimensional and three-dimensional CIT. Results from the zero-dimensional and three-dimensional simulations were already reported above, but are repeated here for comparison purposes. The two-dimensional case is computed in the same manner as in three dimensions, but with one grid point in z and a length corresponding to the mean initial inter-particle spacing.

As shown in Figure 2.8, the two-dimensional case exhibits the most variance in temperature and mass fractions and reacts more slowly compared to the other two cases. This is consistent with previous findings reported in the literature. Capecelatro et al. [40] showed that two-dimensional simulations over-predict clustering, and later Guo and Capecelatro [56] showed clustering delays heat transfer. Therefore, two-dimensional simulations will systematically over-predict spatial heterogeneity in the

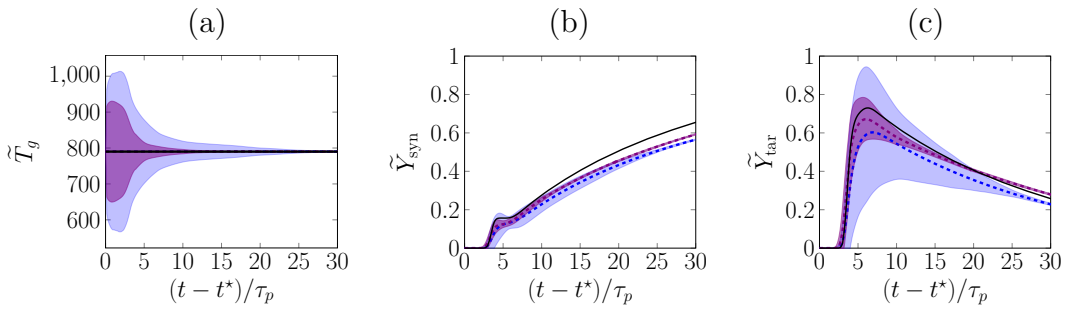


Figure 2.8: Comparison between the zero-dimensioanl (black line) configuration and two-dimensional CIT (blue dashed line) and three-dimensional CIT (violet dashed line). Lines depict mean values and the shaded regions denote ±3 times the standard deviation.

Component	0D	2D	3D
Tar	0.73	0.60	0.67
Syngas	0.16	0.14	0.15
Char	0.33	0.29	0.38

Table 2.3: Mean values of tar, syngas, and char after $6.25\tau_p$, representative of the typical residence time in a riser.

riser and under estimate product yields. A summary of predicted yields for tar, syngas, and char after $6.25\tau_p$ (corresponding to a typical residence time in a riser) is provided in Table 3. Except for char, the zero-dimensional simulation systematically over-predicts product yields compared to the three-dimensional case, and the two-dimensional simulation systematically under-predicts these values (due to increased heterogeneity). Interestingly, the conversion of activated biomass to char is greatest in three dimensions.

2.4 Conclusions

In this work, highly-resolved Eulerian-Lagrangian simulations were performed to model biomass conversion in the fully-developed region of a riser. Strong momentum coupling between the phases leads to the spontaneous generation of clusters that are largely made up of sand. Due to the reduced drag felt by particles within clusters,

both sand and biomass reach a mean settling velocity approximately two times their terminal velocities. At early stages of the conversion process, biomass is initially cooler than the surrounding gas and predominantly located within clusters. The gas phase is cooled relatively fast within clusters and remains hot in regions void of particles. Despite this, tar and syngas are initially generated within clusters due to the large quantity of biomass particles in these regions. They are then advected and diffused away from clusters and secondary reactions take place. After sufficient time, the concentration of tar and syngas becomes spatially uniform. Spatial heterogeneity in CIT results in non-negligible reductions in syngas and tar yields. Furthermore, two-dimensional simulations over-predict the level of clustering, and was found to under-predict syngas and tar yields. In conclusion, three-dimensional multiphase dynamic effects have a direct consequence on transport processes and product yield in biofuel reactors and should be considered in reduced-order models typically employed in industrial processes. Specifically, models that solve for averaged quantities (e.g., Reynolds-averaged Navier–Stokes) require closure for a number of scalar fluxes that contain correlations between fluctuating values [115]. Our previous work has shown that the effect of clusters on heat and mass transfer are contained in the volume fraction-mass fraction covariance [40, 71, 56]. However, mean gradient diffusion models typically used in single-phase turbulence [116] fail to predict these terms. Using machine learning techniques to determine the specific functional form of these dependencies is the subject of future work.

CHAPTER III

Developing a Modeling Method: Sparse Regression with Embedded Form Invariance

A data-driven framework for formulation of closures of the Reynolds-Average Navier–Stokes (RANS) equations is presented. In recent years, the scientific community has turned to machine learning techniques to translate data into improved RANS closures. While the body of work in this area has primarily leveraged Neural Networks (NNs), we alternately leverage a sparse regression framework. This methodology has two important properties: (1) The resultant model is in a closed, algebraic form, allowing for direct physical inferences to be drawn and naive integration into existing computational fluid dynamics solvers, and (2) Galilean invariance can be guaranteed by thoughtful tailoring of the feature space. Our approach is demonstrated for two classes of flows: homogeneous free shear turbulence and turbulent flow over a wavy wall. The model learned based upon the wavy wall configuration is then validated against flow over a backward-facing step. This work demonstrates similar performance to that of modern NNs but with the added benefits of interpretability, increased ease-of-use and dissemination, and robustness to sparse and noisy training datasets.

3.1 Introduction

Simulation frameworks based on the Reynolds-Averaged Navier–Stokes (RANS) equations [14, 117, 118, 119] have been the most widely-used tool in industrial and large-scale applications of turbulent flows for the last several decades [11] and will remain to be a central tool for guiding design decisions well into the coming decades [120, 121]. This is primarily driven by the wide range of length- and time-scales associated with turbulent flows of interest. Because of this, direct numerical simulations (DNS) that fully resolve all relevant scales are prohibitively costly. Instead, the RANS equations solve for mean flow quantities that are then used to assess global flow features of interest. A principal challenge associated with RANS is accurate modeling of the unresolved terms, which are denoted 'unclosed' because they are not completely specified in terms of the unknowns (e.g., mean velocity, pressure, etc.).

With the rise of computational power and the accessibility of large, highly resolved datasets, the community has turned to machine learning techniques in recent years to distill this wealth of information into improved RANS models. As a consequence of the interest and prevalence of the use of machine learning in the turbulence modeling community, several thoughtful and thorough reviews have been published and the authors refer the interested reader to several of these works [see, e.g. 122, 21, 26].

Numerous studies in recent years approach the RANS closure problem by leveraging a Neural Network (NN)-based framework. Ling et al. [28] used an invariant tensor basis integrated into a NN to model the Reynolds stress anisotropy tensor for turbulent duct flow as well as flow over a wavy wall. Galilean invariance, a critical model property, was ensured by nature of the invariant tensor basis as the inputs to the NN and demonstrated excellent agreement with DNS data as compared to traditional (linear and quadratic eddy viscosity) models. Following this work, many others have implemented similar strategies and employed a similar basis technique for ensuring invariance. Of these works, many have used flow through a periodically constricted

channel or backward facing step as challenging tests of new modeling methodologies as these flows exhibit massive separation. This is notoriously difficult to accurately capture with traditional RANS closures [123]. A large body of works using NNs as the data-driven methodology to formulate closure models have demonstrated promising success [124, 125, 126, 28, 30, 25].

Despite the demonstration of improved model performance, models based upon NNs have an important drawback. Due to the nature of the algorithm at the heart of NNs, the resultant model acts as a 'black box' and cannot be expressed in a compact, algebraic form. This compromises interpretability, introduces difficulty in disseminating the learned model with end users and industries, and increases computational cost of the model in the context of a RANS solver (as compared with traditional algebraic closures). Further, a large number of NN approaches attempt to augment or correct existing models. However, this approach breaks down for more complex turbulent flows, such as disperse two-phase flows [39, 40, 41, 127, 6] or turbulent combustion [43, 44], in which the fundamental assumption of an energy cascade breaks down due to production at the smallest scales. In these cases, existing closures adopted from single-phase flows are not appropriate, which precludes an augmentation modeling approach. For these reasons, the present study proposes an alternate method that allows for the development of physics-based, compact algebraic closures, thus affording interpretability, transportability and efficiency.

Several studies have taken alternate approaches to NNs using symbolic methods in order to arrive at closed form, algebraic models. Gene Expression Programming [31, 32, 33, 34, 35] and random forest regression [128, 37] have become increasingly popular methodologies. The early success of these works serves as motivation for the present work in which we present a methodology based upon sparse regression as an alternative to NNs for developing new RANS closures, with emphasis on the following key benefits,

- *Interpretability*: Sparse regression produces an algebraic model with a limited number of terms, resulting in improved interpretability of underlying physics and better prediction of model behavior and stability outside the scope of training.

- *Galilean invariance*: By careful construction of the feature space and structuring of the optimization cost functional, Galilean invariance of the resultant model is ensured.

- *Efficiency*: Models developed using sparse regression are built using physics-based, functional terms and identifies a *subset* of these terms that are most important for capturing physics. This is fundamentally different from a naive curve fit in which all possible terms are included in the model. Thus, forward computations are necessarily fewer compared to other techniques, such as NNs, which postulate a full rank model. As a consequence of determining a simpler model form, we also demonstrate comparable performance to NNs which were trained on larger datasets. Further, the resultant model is algebraic, making for a lighter and more efficient integration with existing solvers.

Beyond developing the methodology, its utility is demonstrated on two canonical cases: homogeneous free shear turbulence and turbulence through a periodically constricted channel. Within the context of homogeneous free shear turbulence, the sparse regression methodology is validated using a 'toy' problem in which the training data set is synthetically generated using an existing model. Then, sparse regression is used to recover this existing model. Subsequent cases are based upon DNS data and seek to uncover improved models in comparison with existing closures. Finally, the algorithm is given experimental data for training and this result is compared with those determined using full-field DNS data.

3.2 Methodology

The sparse regression approach expands upon the data-driven technique presented in Brunton et al. [38] for using temporally evolving data to 'discover' nonlinear, dynamical systems. Rather than uncovering governing equations, this method is employed to identify robust, data-driven closure models. In this section, the Sparse Identification of Nonlinear Dynamics (SINDy) framework [38] is built upon by adapting it for the RANS closure problem and embedding invariance–a key property of any candidate RANS model.

It is first postulated that a tensor quantity of interest, $\mathbb{D}$, can be characterized by the linear combination of an invariant tensor basis, represented as $\mathbb{T}$, premultiplied by optimal coefficients, represented as $\hat{\boldsymbol{\beta}}$,

$$\mathbb{D} = \mathbb{T}\hat{\boldsymbol{\beta}}. \tag{3.1}$$

Using this postulated form of the model, the following objective function is minimized in order to determine the optimal coefficient vector, according to

$$\hat{\boldsymbol{\beta}} = \min_{\boldsymbol{\beta}} ||\mathbb{D} - \mathbb{T}\boldsymbol{\beta}||_2^2 + \lambda||\boldsymbol{\beta}||_1, \tag{3.2}$$

where β represents intermediary realizations of the coefficient vector which may not necessarily be the optimal coefficient vector, $\hat{\beta}$. Here, the L-2 and the L-1 norms are denoted by $||\cdot||_2^2$ and $||\cdot||_1$, respectively. The first term in the objective function is ordinary least squares, which regresses the coefficient vector to the trusted data, and the second term is a sparsity-inducing penalty on the coefficients. By choice of the L-1 penalty, the minimization of the objective function performs model selection by inducing sparsity (e.g., several of the terms of $\hat{\boldsymbol{\beta}}$ are identically zero, indicating that the associated term in the invariant basis, $\mathbb{T}$, is not important. The interested

reader can refer to [129, 130, 131] for further information.) Minimization of the cost function is performed using the open source iterative algorithm presented in Brunton et al. [38].

In order to obtain a model that is both compact and frame-invariant, consideration must be given to the construction of the trusted data vector, $\mathbb{D}$, and the invariant basis, $\mathbb{T}$. For compactness, $\mathbb{D}$, and as a consequence $\boldsymbol{\beta}$, are restricted to column vectors. This ensures that the coefficients for each term in the model is a scalar, which guarantees the same model form regardless of orientation (e.g., if the coefficients were vectors or tensors, this would embed directionality into the coefficients and thereby enslave the model to the orientation in which it was learned).

In this work, $\mathbb{D}$ is assembled by first assessing the symmetry of the problem. All nonzero, unique entries in the trusted data tensor are concatenated into a column vector. For example, as seen in Figure 3.1, if $\mathcal{D}$ is symmetric in the y– and z– directions and the only anisotropic contribution is in the x–y direction, then the full tensor is represented as $[\mathcal{D}_{11}, \mathcal{D}_{12}, \mathcal{D}_{22}]^{\mathrm{T}}$. For each realization (e.g., in time) and for each configuration under consideration, these column vectors are vertically concatenated.

Finally, form (Galilean) invariance in the resultant model is guaranteed by assembling $\mathbb{T}$ from an invariant tensor basis. The basis is crafted by using dimensional analysis to determine the relevant known tensor quantities that fully describe the physics under study. These tensors are then used to assemble a minimal integrity basis (see, e.g., Pope [132], Speziale et al. [133], Ling et al. [28]), using the following arguments:

1. Any tensor can be represented by an infinite tensor sum of the form:

$$\mathcal{D}_{ij} = \sum_{n=1}^{\infty} G^{(n)} \mathcal{T}_{ij}^{(n)},$$

where G are coefficients that in the general sense may be functions of the in-

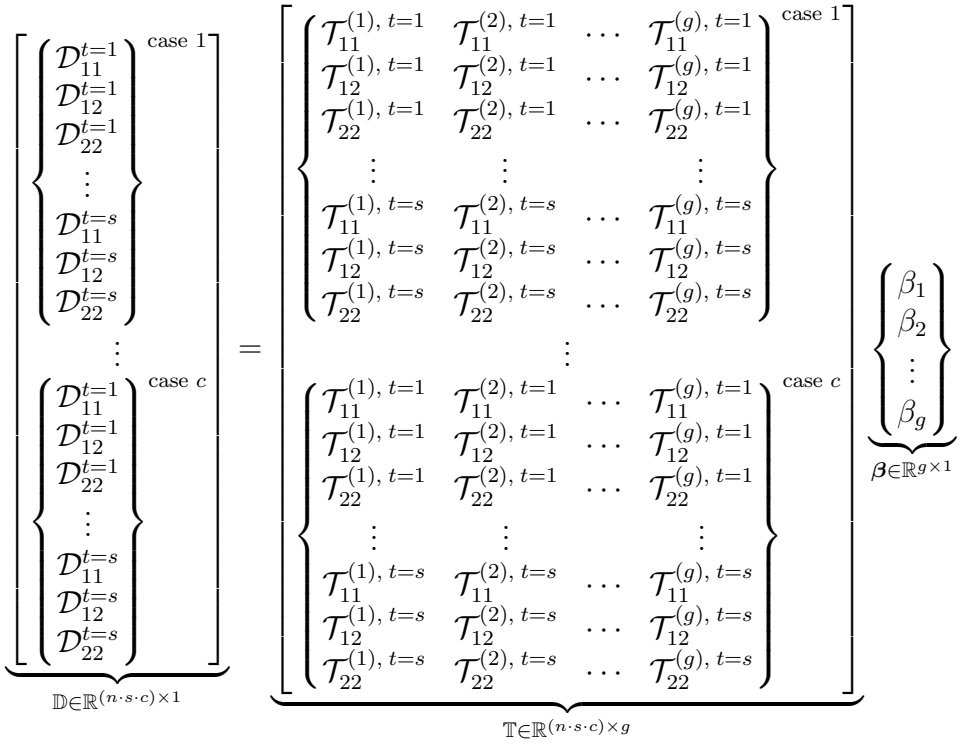


Figure 3.1: The postulated model takes the form $\mathbb{D}_i = \mathbb{T}_{ij}\beta_j$, where $\mathbb{D}$ contains the observed data spanning over c cases, each with s realizations in time. Further, $\mathbb{T}$ spans an invariant tensor basis of dimension g evaluated at each of the samples s for each case c.

variants of the tensor basis $\mathcal{T}_{ij}^{(n)}$.

2. In some cases, the Cayley-Hamilton theorem can be leveraged to reduce the infinite tensor sum to a finite sum that still exactly represents the infinite sum. In cases where this is not possible, the basis is truncated once model improvement stagnates.

Once the invariant basis is determined, the matrix $\mathbb{T}$ is assembled using the same convention as for $\mathbb{D}$. Then, the sparsity parameter λ is adjusted until acceptable model error and sparsity are reached, noting that $\lambda = 0$ is Ordinary Least Squares, and increasing λ results in an increasingly sparse coefficient vector, $\hat{\boldsymbol{\beta}}$.

In the following sections, the sparse regression methodology is applied to two canonical cases of increasing complexity. First, homogeneous free shear turbulence is considered. As an initial proof-of-concept, a synthetic dataset is generated using a known model and the sparse regression methodology is used to recover that model. Next, DNS is used to generate the trusted datasets for the same physical configuration and sparse regression is used to uncover alternate models to those traditionally used. Last, turbulence through a periodically constricted channel is considered. This test case has the additional complexities of being a statistically two-dimensional flow (as compared to homogeneous free shear turbulence being statistically one-dimensional in time) as well as giving rise to flow separation. The model learned by sparse regression is compared with the standard Linear Eddy Viscosity Model (LEVM) and is then evaluated outside the scope of its training for the periodically constricted channel at a higher Reynolds number as well as for flow over a backward facing step. Finally, sparse regression is employed to train a model using experimental data for the periodically constricted channel. This learned model is then compared and contrasted with the model learned using the full-field DNS data.

3.3 Case studies

Herein, Reynolds decomposition is denoted by angled brackets, $\langle\cdot\rangle$, given for the velocity vector by $u_i(x_i,t) = \langle u_i(x_i,t)\rangle + u_i'(x_i,t)$, where $u_i(x_i,t)$ is the field quantity for velocity written in Einstein notation, x_i is the location, and t is time with $\langle u_i(x_i,t)\rangle$ and $u_i'(x_i,t)$ denoting the mean (which may be spatial, temporal or both) and the fluctuating portions of the velocity, respectively. Applying Reynolds averaging to the incompressible Navier–Stokes equations yields the RANS equations,

$$\frac{\partial\langle u_i\rangle}{\partial x_i} = 0 \tag{3.3}$$

$$\frac{\partial\langle u_i\rangle}{\partial t} + \langle u_k\rangle\frac{\partial\langle u_i\rangle}{\partial x_k} = -\frac{1}{\rho}\frac{\partial\langle p\rangle}{\partial x_i} + \frac{\partial}{\partial x_j}\left[\nu\left(\frac{\partial\langle u_i\rangle}{\partial x_j} + \frac{\partial\langle u_j\rangle}{\partial x_i}\right) - \langle u_i'u_j'\rangle\right]. \tag{3.4}$$

It is notable that the Reynolds averaging process yields a Reynolds stress term, $\langle u_i'u_j'\rangle$, which requires closure.

The strategy for closure of the Reynolds stress term generally falls into two categories: (1) an algebraic closure or (2) the inclusion of a transport equation for the Reynolds stresses. In this work, two flows serve as case studies for the implementation of the methodology described in Section 3.2. The first case study (homogeneous free shear turbulence) will develop closures in the form of transport of the Reynolds stresses and the second (turbulence in a periodically constricted channel) will consider algebraic closure.

3.3.1 Homogeneous free shear turbulence

3.3.1.1 Problem statement

The flow configuration under consideration in this section is homogeneous free shear turbulence, in which an unbounded, three-dimensional fluid volume is subjected to a mean-velocity gradient that generates and sustains turbulence. After sufficient

time, the Reynolds stresses reach a 'self-similar' state, characterized by the anisotropy of the Reynolds stresses reaching stationarity in time (e.g., $\frac{d}{dt}\left(\langle u_i' u_j' \rangle / k\right) = 0$ with $k = \langle u_k' u_k' \rangle$ the turbulent kinetic energy (TKE)). Consequently, Reynolds-averaged quantities are statistically one-dimensional (i.e., they depend only on time). It is this 'self-similar' behavior that is of specific interest in formulating an improved RANS closure.

As previously described, the Reynolds stresses in the RANS equations (Eq. (3.4)) require closure. In this example, we consider the transport of the Reynolds stresses, which are given exactly as

$$\begin{aligned}\frac{D\langle u_i' u_j' \rangle}{Dt} = &\underbrace{-\left[\langle u_j' u_k' \rangle \frac{\partial \langle u_i \rangle}{\partial x_k} + \langle u_i' u_k' \rangle \frac{\partial \langle u_j \rangle}{\partial x_k}\right]}_{\text{production, } \mathcal{P}_{ij}} \underbrace{- 2\nu \left\langle \frac{\partial u_i'}{\partial x_k} \frac{\partial u_j'}{\partial x_k} \right\rangle}_{\text{dissipation, } \varepsilon_{ij}} + \underbrace{\left\langle \frac{p}{\rho}\left(\frac{\partial u_i'}{\partial x_j} + \frac{\partial u_j'}{\partial x_i}\right)\right\rangle}_{\text{redistribution, } \mathcal{R}_{ij}} \\ &\underbrace{- \frac{\partial}{\partial x_k}\langle u_i' u_j' u_k' \rangle}_{\text{turbulent convection}} + \underbrace{\nu \frac{\partial^2 \langle u_i' u_j' \rangle}{\partial x_k^2}}_{\text{viscous diffusion}} \underbrace{- \frac{\partial}{\partial x_k}\left(\frac{\langle u_i' p' \rangle}{\rho}\delta_{jk} + \frac{\langle u_j' p' \rangle}{\rho}\delta_{ik}\right)}_{\text{pressure transport}}\end{aligned} \tag{3.5}$$

where D/Dt denotes the material derivative, ν is the kinematic viscosity, and ρ and p denote fluid density and pressure, respectively. In the case of homogeneous free shear turbulence, the domain is spatially homogeneous and consequently, spatial gradients of mean quantities are null. Thus, the transport of Reynolds stresses is reduced to

$$\frac{d\langle u_i' u_j' \rangle}{dt} = \underbrace{-\left[\langle u_j' u_k' \rangle \Gamma_{ik} + \langle u_i' u_k' \rangle \Gamma_{jk}\right]}_{\text{production, } \mathcal{P}_{ij}} \underbrace{- 2\nu \left\langle \frac{\partial u_i'}{\partial x_k} \frac{\partial u_j'}{\partial x_k} \right\rangle}_{\text{dissipation, } \varepsilon_{ij}} + \underbrace{\left\langle \frac{p}{\rho}\left(\frac{\partial u_i'}{\partial x_j} + \frac{\partial u_j'}{\partial x_i}\right)\right\rangle}_{\text{redistribution, } \mathcal{R}_{ij}}, \tag{3.6}$$

where the shear rate tensor is given as $\Gamma_{ij} = \partial \langle u_i \rangle / \partial x_j$, Here, the production term is closed, however the dissipation and redistribution tensors both require closure. In this work, new modeling efforts are directed toward the redistribution tensor and the dissipation tensor is closed using the standard transport equation proposed by

Hanjalic and Launder [134],

$$\frac{\partial \varepsilon}{\partial t} = C_{\varepsilon 1}\frac{\mathcal{P}\varepsilon}{k} - C_{\varepsilon 2}\frac{\varepsilon^2}{k}, \tag{3.7}$$

where $\mathcal{P} = \mathrm{tr}\left(\mathcal{P}_{ij}\right)/2$ and model constants are given by $[C_{\varepsilon 1}, C_{\varepsilon 2}] = [1.44, 1.92]$ [135].

3.3.1.2 Proof-of-concept: a synthetic dataset

As an initial proof-of-concept for the sparse regression methodology described in Section 3.2, a set of data is generated using a well-established closure for the redistribution tensor with the goal of recovering the known model. The closure utilized to generate the synthetic dataset was proposed by Launder et al. [18] and is known as the LRR-IP model,

$$\mathcal{R}_{ij} = -C_R\frac{\varepsilon}{k}\left(\langle u_i' u_j'\rangle - \frac{2}{3}k\delta_{ij}\right) - C_2\left(\mathcal{P}_{ij} - \frac{2}{3}\mathcal{P}\delta_{ij}\right), \tag{3.8}$$

where the constants are given as $[C_R, C_2] = [1.8, 0.6]$ [18]. This closure, embedded in the transport equation for the Reynolds stresses in Eq. (3.6) and the transport equation for dissipation given in Eq. (3.7) are solved for three shear rates ($\Gamma = \Gamma_{12} = [2.25, 11.24, 20.23]$). This results in one-dimensional (time-dependent) data for the Reynolds stresses for each shear rate.

Given the simple flow configuration, the redistribution tensor can be normalized by the viscous dissipation rate, ε, and characterized by a linear combination of the the following non-dimensionalized, mean flow quantities:

$$\begin{aligned}
&\text{(1) Anisotropic stress tensor} && b_{ij} = && \tfrac{\langle u_i' u_j'\rangle}{2k} - \tfrac{1}{3}\delta_{ij} \\
&\text{(2) Mean rotation rate tensor} && \hat{R}_{ij} = && \tfrac{1}{2}\tfrac{k}{\varepsilon}\left(\tfrac{\partial \langle u_i\rangle}{\partial x_j} - \tfrac{\partial \langle u_j\rangle}{\partial x_i}\right) \\
&\text{(3) Mean shear rate tensor} && \hat{S}_{ij} = && \tfrac{1}{2}\tfrac{k}{\varepsilon}\left(\tfrac{\partial \langle u_i\rangle}{\partial x_j} + \tfrac{\partial \langle u_j\rangle}{\partial x_i}\right)
\end{aligned}$$

such that,

$$\mathcal{R}_{ij} = \varepsilon \boldsymbol{f}(b_{ij}, \hat{R}_{ij}, \hat{S}_{ij}), \tag{3.9}$$

where $\boldsymbol{f}$ is a *form invariant* tensor-valued function, which, due to the linearity in $b_{ij}, \hat{R}_{ij}$, and $\hat{S}_{ij}$ automatically satisfies,

$$\boldsymbol{Q}\boldsymbol{f}(b_{ij}, \hat{R}_{ij}, \hat{S}_{ij})\boldsymbol{Q}^{\mathrm{T}} = \boldsymbol{f}(\boldsymbol{Q}b_{ij}\boldsymbol{Q}^{\mathrm{T}}, \boldsymbol{Q}\hat{R}_{ij}\boldsymbol{Q}^{\mathrm{T}}, \boldsymbol{Q}\hat{S}_{ij}\boldsymbol{Q}^{\mathrm{T}}). \tag{3.10}$$

Here, $\boldsymbol{Q}$ is a Galilean rotation matrix, e.g., $\boldsymbol{Q}\boldsymbol{Q}^{\mathrm{T}} = \boldsymbol{Q}^{\mathrm{T}}\boldsymbol{Q} = \boldsymbol{I}$ (where $\boldsymbol{I}$ is the identity tensor) and $\det \boldsymbol{Q} = \pm 1$.

These bases have been extensively used in the literature (see e.g, [133, 136] for their derivation) and their usage is restricted to modeling equilibrium regimes (rather than the transient period). To briefly summarize, because the redistribution tensor, $\mathcal{R}_{ij}$, is symmetric and deviatoric, and its dependence on each of the bases is linear, each basis tensor must also satisfy these same properties. Further, the constraint of form invariance under coordinate transformation stipulates that $\boldsymbol{f}$ must be an isotropic function of its arguments (i.e., b_{ij}, $\hat{R}_{ij}$, and $\hat{S}_{ij}$). Using these constraints guides the formulation of the minimal integrity basis [133] as shown in the leftmost column of Table 3.1.

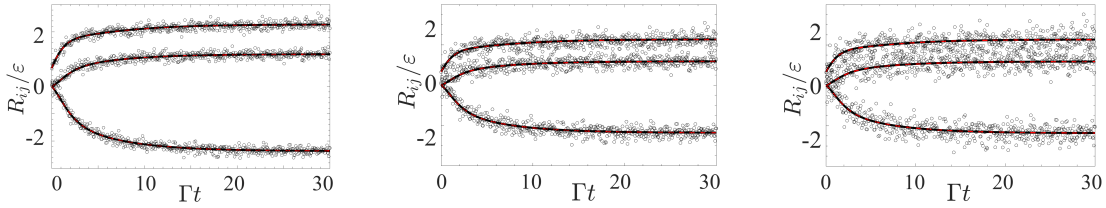
Using the data generated by solving Eqs. (3.6)–(3.14), the basis tensors are computed and the redistribution tensor is populated by taking the time derivative of the Reynolds stresses (using a 6th-order central difference scheme) and solving Eq. (3.6) for the redistribution tensor. Then, these quantities are assembled into $\mathbb{D}$ and $\mathbb{T}$ as described in Section 3.2. Note that since we are interested in modeling the self-similar regime, only data from this region is used for training and for assessing model error. After $\mathbb{D}$ and $\mathbb{T}$ are assembled, the cost functional defined by Eq. (3.2) is optimized for decreasing values of λ until reduction in model error is no longer achieved.

As shown in Table 3.1, the methodology exactly returns the LRR-IP model used to generate the dataset. In order to systematically challenge the robustness of the algorithm, *a posteriori*, artificial noise was added to the synthetic data set. The syn-

thetic noise was normally distributed about the mean of the synthetic data, denoted as $\mathcal{N}(\mu, \sigma)$, where σ is the standard deviation that is prescribed in terms of percentage of the mean value, μ. We consider $\sigma = 0.1, 0.2$, and 0.3μ. In each case, λ was reduced until the model error plateaued. Even in the case of the noisiest data provided, the learned model deviated from the expected LRR-IP model by only 2.3%, where error is defined by the$L$-2 norm

$$\epsilon = \frac{||\mathbb{D} - \mathbb{T}\beta||_2}{||\mathbb{D}||_2}. \tag{3.11}$$

This level of performance indicates the sparse regression methodology is robust to substantial noise in the training data without compromising the accuracy in learning the underlying physics. This is further demonstrated in Figs. 3.2 (a)–3.2 (c) where the models learned from the noisy data are shown against the LRR-IP model and the artificially noisy data. In all three cases, the learned model accurately describes the behavior of the LRR-IP model despite small amounts of error in the coefficients.



(a) $\mathcal{N}(\mu, 0.1\mu)$ (b) $\mathcal{N}(\mu, 0.2\mu)$ (b) $\mathcal{N}(\mu, 0.3\mu)$

Figure 3.2: Comparison between prescribed LRR-IP model (—) with the learned models (- - -) and artificially noisy data (∘), for $\Gamma = 2.25$.

3.3.1.3 DNS-generated data: Can sparse regression improve upon existing models?

Here, the same physical configuration is considered, albeit with the trusted data generated using DNS (see Figure 3.3). The goal of using DNS-generated training data is to understand how the learned models differ from classically used models.

To generate the DNS datasets, NGA [85], a fully conservative, low-Mach num-

order in b_{ij}	$\mathcal{T}^{(n)}$	LRR-IP	Sparse Regression $\lambda = 0.1$ $\mathcal{N} = 0$	$\lambda = 0.1$ $\mathcal{N}(\mu, 0.1\mu)$	$\lambda = 0.1$ $\mathcal{N}(\mu, 0.2\mu)$	$\lambda = 0.5$ $\mathcal{N}(\mu, 0.3\mu)$
0	S_{ij}	0.8	0.8	0.8010	0.8020	0.803
	b_{ij}	-3.6	-3.6	-3.5761	-3.5522	-3.5282
1	$R_{il}b_{lj} + R_{jl}b_{li}$	1.2	1.2	1.2003	1.2007	1.2010
	$S_{il}b_{lj} + S_{jl}b_{li} - \frac{2}{3}S_{lm}b_{ml}\delta_{ij}$	1.2	1.2	1.2018	1.2036	1.2054
	$b_{ij}^2 - \frac{1}{3}b_{ll}^2\delta_{ij}$	0	0	0	0	0
2	$S_{il}b_{lj}^2 + S_{jl}b_{li}^2 - \frac{2}{3}S_{lm}b_{ml}^2\delta_{ij}$	0	0	0	0	0
	$R_{il}b_{lj}^2 + R_{jl}b_{li}^2$	0	0	0	0	0
3	$b_{ik}^2R_{kp}b_{pj} - b_{il}R_{lk}b_{kj}^2$	0	0	0	0	0
	$\epsilon^{b_{ij}}$	–	*0.0*	*0.0076*	*0.015*	*0.023*

Table 3.1: Summary of model forms and associated error in the self-similar region of homogeneous free shear turbulence, with increasing amounts of artificial noise added to the synthetic dataset.

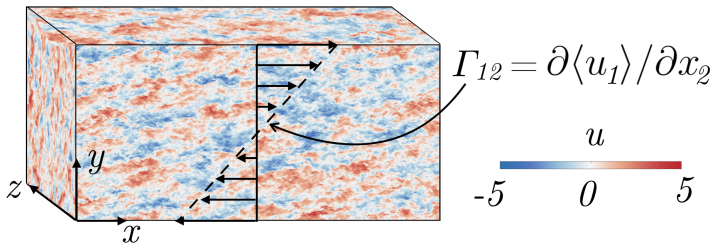


Figure 3.3: Snapshot of the instantaneous velocity field in DNS homogeneous free shear turbulence at $\mathcal{S} = 11.2$.

ber finite volume solver is used. Details on this code have been previously desrcibed in Section2.2.3. Shear periodic boundary conditions are enforced using the recently developed algorithm of Kasbaoui et al. [137]. Turbulence in the domain is initialized using spectral methods in order to ensure consistency with Kolmogorov's '-5/3' spectrum [138, 139].

Five cases are simulated for non-dimensional shear rates $\mathcal{S} = 2\Gamma k_0/\varepsilon_0 = (2.3, 6.6, 11.2, 13.2, 20.2)$ on a grid of size $1024 \times 512 \times 512$, corresponding to a domain size of $2\pi \times \pi \times \pi$. Here, k_0 and ε_0 denote the initial values of TKE and dissipation, respectively. The grid resolution ensures that the flow captures the dissipative scales. Each case is simulated to a non-dimensional time of $\Gamma t \approx 25 - 30$ to ensure sufficient data in the self-similar region is captured. Of the five datasets, three are selected as training sets ($\mathcal{S} = 2.3, 11.2, 20.2$) from which a new model is learned. The remaining two datasets ($\mathcal{S} = 6.6, 13.2$) serve as validation sets in order to determine the optimal value of λ and therefore the optimal learned model.

In the same fashion as was described for the synthetic dataset, the DNS data is organized into $\mathbb{D}$ and the $\mathbb{T}$, and the cost functional is optimized for decreasing values of the sparsity parameter λ until all terms are populated in the learned model. The resulting models from this procedure are shown in Table 3.2. As λ is decreased, additional terms are included in the learned model and the coefficients adjust accordingly. The four learned models are compared against existing models, the Rotta

[140], LRR-IP and LRR-QI models [18] (Eqs. (3.12)–(3.15)), written in terms of the basis tensors as,

$$\mathcal{R}_{ij}^{\text{Rotta}} = -2C_R b_{ij}, \tag{3.12}$$

$$\mathcal{R}_{ij}^{\text{LRR-IP}} = -2C_R b_{ij} + \frac{4}{3}C_2 S_{ij} + 2C_2\left(R_{il}b_{lj} + R_{jl}b_{li}\right) + 2C_2\left(S_{il}b_{lj} + S_{jl}b_{li} - \frac{2}{3}S_{lm}b_{ml}\delta_{ij}\right), \tag{3.13}$$

$$\mathcal{R}_{ij}^{\text{LRR-QI}} = -2C_R' b_{ij} + \frac{4}{5}S_{ij} + \frac{2}{11}\left(10 - 7C_2'\right)\left(R_{il}b_{lj} + R_{jl}b_{li}\right) + \frac{6}{11}\left(2 + 3C_2'\right)\left(S_{il}b_{lj} + S_{jl}b_{li} - \frac{2}{3}S_{lm}b_{ml}\delta_{ij}\right). \tag{3.14}$$

Here, the coefficients are given as $[C_R, C_2] = [1.8, 0.6]$ and $[C_R', C_2'] = [1.5, 0.4]$. The Rotta model assumes a linear relationship with the anisotropy tensor and thus models a linear return to isotropy. In contrast, the LRR-IP model includes nonlinear terms that are important for characterizing homogeneous *anisotropic* turbulence. In comparing these three models with four learned models of increasing complexity, it is observed that the least complex learned model, corresponding to $\lambda = 0.75$, already shows marked improvement over the highest performing existing models and reduces error in the anisotropic stress tensor from 26% to 9%.

Shown in Figure 3.4, as λ is decreased and terms are added to the learned model, the normalized coefficients, $\tilde{\beta}_i = \beta_i / \left(\max \beta_i\right)$, change to accommodate contributions from additional terms. In Figure 3.4 (a), the sparse regression methodology is employed on a basis that is restricted to up to first order in b_{ij}. The basis is then expanded to second and third order terms in b_{ij} in Figures. 3.4 (b) and 3.4 (c), respectively. In each instance, the most prominent coefficient remains the largest contribution to the learned model, though its contribution decreases as subsequent terms are added. The order of prominence of the lesser contributing terms does not remained fixed once the number of terms in the model grows. This behavior has an insignificant effect on model performance and sensitivity and serves to demonstrate

order in b_{ij}	$\mathcal{T}^{(n)}$	Rotta	LRR-IP	LRR-QI	Sparse Regression $\lambda = 0.75$	$\lambda = 0.6$	$\lambda = 0.5$	$\lambda = 0$
0	S_{ij}	0	0.8	0.8	1.01	1.01	0.98	0.98
1	b_{ij}	-3.6	-3.6	-3.0	1.27	1.31	1.45	1.46
	$R_{il}b_{lj} + R_{jl}b_{li}$	0	1.2	1.31	1.53	1.56	1.49	1.48
	$S_{il}b_{lj} + S_{jl}b_{li} - \frac{2}{3}S_{lm}b_{ml}\delta_{ij}$	0	1.2	1.74	1.73	1.71	1.79	1.78
2	$b_{ij}^2 - \frac{1}{3}b_{ll}^2\delta_{ij}$	0	0	0	5.22	4.64	7.02	6.71
	$S_{il}b_{lj}^2 + S_{jl}b_{li}^2 - \frac{2}{3}S_{lm}b_{ml}^2\delta_{ij}$	0	0	0	0	0	0.57	0.56
	$R_{il}b_{lj}^2 + R_{jl}b_{li}^2$	0	0	0	0	0	0	0.13
3	$b_{ik}^2R_{kp}b_{pj} - b_{il}R_{lk}b_{kj}^2$	0	0	0	0	-0.65	2.08	2.45
	training error, $\epsilon_{\text{train}}^{\boldsymbol{b}}$	0.68	0.26	0.26	0.090	0.092	0.078	0.073
	testing error, $\epsilon_{\text{test}}^{\boldsymbol{b}}$				0.086	0.089	0.070	0.078

Table 3.2: Summary of learned and existing models and associated error in the self-similar region for homogeneous free shear turbulence.

the relative lesser importance of these terms to the overall model performance as compared with the terms with larger contributions. Finally, it is observed in Figure 3.4 that the dominant coefficient stagnates beyond a five term model. This mirrors the reduction in overall model error as shown in Table 3.2. The ability to identify which basis tensors are most important for modeling the flow and their relative sensitivity therein, is a unique benefit of the sparse regression methodology which allows for interpretability of the relationship between model form and flow physics.

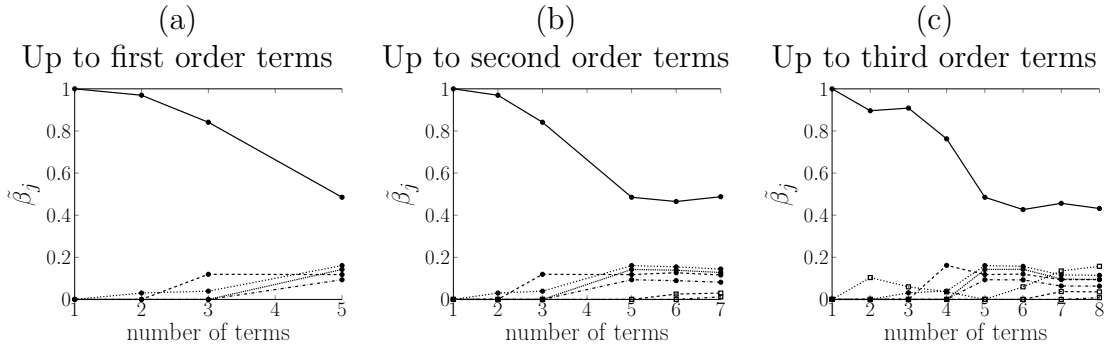


Figure 3.4: As the number of terms in the model increases (by decreasing λ), terms that are most important for capturing key redistribution physics arise in the sparsest models and persist with prominent coefficients as terms are added.

A comparison of the training and validation errors give the clearest indication of when a learned model begins to exhibit symptoms of over-fitting (see Figure 3.5). While the learned closure predicts the redistribution tensor, $\mathcal{R}_{ij}$, the ultimate goal is to improve performance in predicting anisotropy in the Reynolds stresses, b_{ij}, making both measures of error relevant to assessing learned models. As shown in Figure 3.5, the error in $\mathcal{R}_{ij}$ decreases monotonically beyond a two term model, however, we observe that five terms are required for stability in the transport equation for the Reynolds stresses. It is notable that while the LRR-IP and LRR-QI model are stable with four terms, the four term model learned by sparse regression is not. This is likely due to a lack of a stability penalty in the cost function. Adding additional penalties to the cost function in order to enforce model stability is an area of active research.

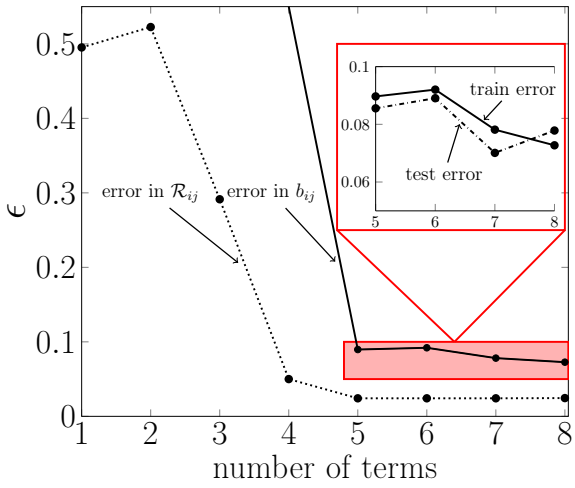


Figure 3.5: Error in $\mathcal{R}_{ij}$ and b_{ij} are shown for models of increasing complexity for homogeneous free shear turbulence. The inset figure delineates validation (test) and training error.

Testing and training errors are also compared in Figure 3.5. As might be expected, the training error generally decreases as terms are added, but beyond seven terms in the learned model, an increase in testing error is observed. This is indicative of over-fitting, thus making the seven term model the ideal model that minimizes model error while maximizing accuracy of the model across different shear rates. As seen in Table 3.2, the ideal learned model reduces error in predicting self-similar behavior by more than half as compared with the LRR-IP or LRR-QI models and more than eight fold as compared with the Rotta model.

In Figure 3.6, the ideal seven-term model is compared with the highest performing existing model, the LRR-QI model. Both are plotted against the DNS values used for training (Figures 3.6 (a)–3.6 (c)) and for validation (Figures 3.6 (d)–3.6 (e)). As previously discussed, it is observed that the learned model accurately captures the self-similar behavior (shown in grey shaded regions) of the normalized Reynolds stresses even in the testing cases which were not seen by the sparse regression method during training.

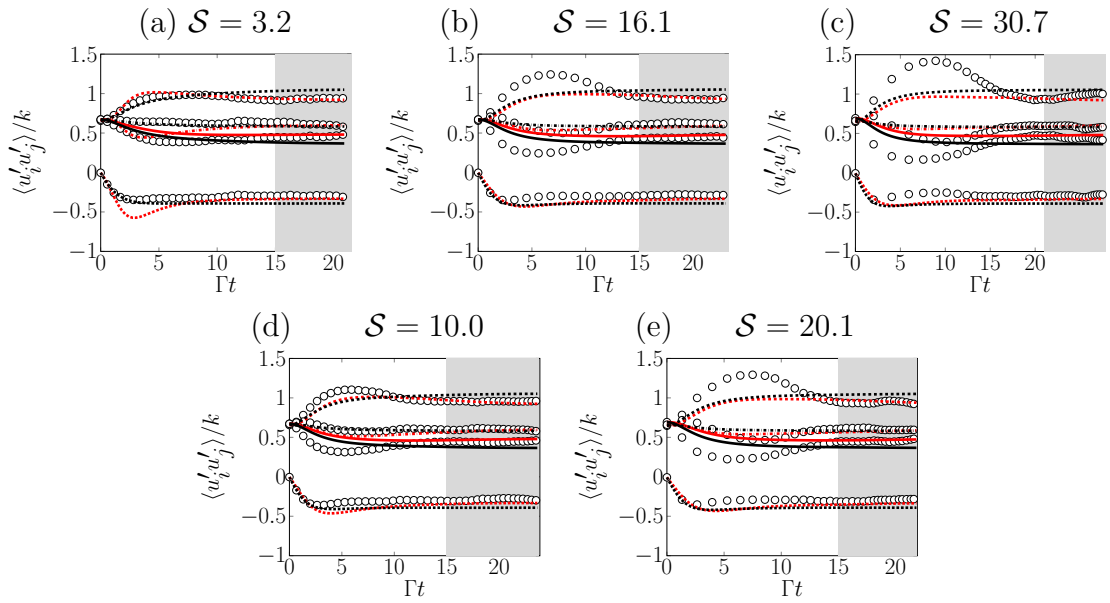


Figure 3.6: Sparse regression (—) produces a more accurate model as compared with the most accurate traditional closure available (LRR-QI, —). DNS data is denoted by open circles and the four lines correspond to the unique components of the normalized Reynolds stresses: $\langle u'u' \rangle$: - - -, $\langle u'v' \rangle$: ·····, $\langle v'v' \rangle$: —, and $\langle w'w' \rangle$ -·-·. The shaded portion denotes the self-similar region of the flow.

3.3.1.4 A note on non-inertial frames of reference

If a *non-inertial* frame is to be considered [141, 142, 136], one would need to modify the normalized, mean rotation rate tensor to include the rotation rate of the frame with respect to an inertial frame ($\mathbf{\Omega}$), i.e. $\hat{\mathbb{R}}_{ij} = \hat{R}_{ij} + \epsilon_{mji}\Omega_m$, where ϵ_{mji} denotes the permutation tensor. Additionally, Coriolis terms, $(\langle u_i u_k \rangle \epsilon_{mkj}\Omega_m + \langle u_j u_k \rangle \epsilon_{mki}\Omega_m)$, must be included in Eqs. (3.4) and (3.6).

3.3.1.5 A note on constant coefficients

The analysis presented above considers the simple case in which model coefficients are constants. As previously discussed, the coefficients are permitted to theoretically depend nonlinearly on the principal invariants of the basis tensors. In the case of homogeneous free shear turbulence, model performance using constant coefficients performs well without the additional complexity of dependency on principal invariants.

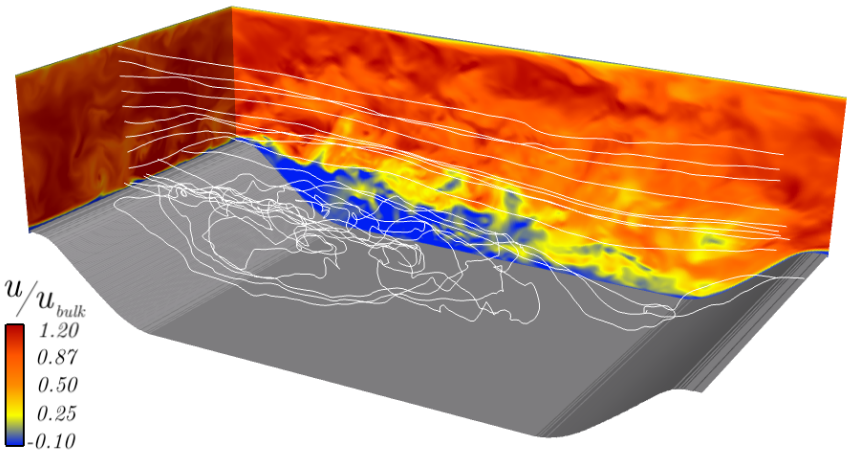


Figure 3.7: Instantaneous streamwise velocity (color), with streamlines originating from $= L_z/2$ (white lines).

However, in situations in which this is not the case, this dependency can be added into $\mathbb{T}$ by postulating functional forms of coefficients and appending them to each of the basis tensors. If there is a physics-based rational for the functional dependency, this process can be prescribed by hand, or if the functional dependence is not in any way constrained, an algorithm such as Gene Expression Programming [31, 32, 33, 34, 143] can be used to analytically determine complex coefficient dependencies on the principal invariants. This strategy is reserved for future work.

3.3.2 Turbulent flow through a periodically constricted channel

3.3.2.1 Problem statement

In this section, we consider the classical case of turbulent flow through a periodically constricted channel as shown in Figure 3.7 and described in Breuer et al. [10]. As discussed in Section 3.3, two main approaches are typically taken when developing closures for the Reynolds stresses. In Section 3.3.1, the transport of the Reynolds stresses was addressed and in this section, algebraic closure of the Reynolds stresses will be developed.

In this strategy, the algebraic closure for the Reynolds stresses depends upon a

model for the anisotropic stress tensor, such that $\langle u_i' u_j' \rangle = 2k(b_{ij} + \frac{1}{3}\delta_{ij})$. Further it has been well established that the model for b_{ij} depends upon $\hat{S}_{ij}$ and $\hat{R}_{ij}$. Recalling from Section 3.3.1 that these quantities are normalized by TKE, k, and dissipation of TKE, ε, this method requires the transport of both k and ε, which are given by

$$\frac{\partial k}{\partial t} + \frac{\partial \left(k u_i\right)}{\partial x_i} = \frac{\partial}{\partial x_j}\left[\left(\nu + \frac{\nu_t}{\sigma_k}\right)\frac{\partial k}{\partial x_j}\right] + \mathcal{P} - \varepsilon, \tag{3.15}$$

$$\frac{\partial \varepsilon}{\partial t} + \frac{\partial \left(\varepsilon u_i\right)}{\partial x_i} = \frac{\partial}{\partial x_j}\left[\left(\nu + \frac{\nu_t}{\sigma_\varepsilon}\right)\frac{\partial \varepsilon}{\partial x_j}\right] + C_{1\varepsilon}\frac{\varepsilon}{k}\mathcal{P} - C_{2\varepsilon}\frac{\varepsilon^2}{k}, \tag{3.16}$$

where $[C_\mu, \sigma_k, \sigma_\varepsilon, C_{1\varepsilon}, C_{2\varepsilon}] = [0.09, 1.00, 1.30, 1.44, 1.92]$. The turbulent viscosity, ν_t, is given as $\nu_t = C_\mu k^2/\varepsilon$ where $C_\mu = 0.09$ [17].

Using these equations along with a model for the anisotropic stress tensor, the RANS equations (Eqs. (3.3) and (3.4)) are closed. The aim of this study is to use sparse regression to develop an improved algebraic closure for the anisotropic stress tensor. As is commonly used in the literature, the configuration under consideration here is turbulent flow through a periodically constricted channel (see Figure 3.7). This flow configuration is particularly challenging because the quantities of interest are statistically 2-D (with dependence on the stream-wise and cross stream directions) and the presence of the constriction generates massive separation in the flow.

The dataset used for training was simulated using NGA, described in Section 3.3.1. The top and bottom walls apply a no-slip boundary condition and the bottom, constricted wall is enforced using a cut-cell immersed boundary method [81]. The geometry for the configuration under study matches the configuration described in Breuer et al. [10] with uniform grid spacing discretized by $[N_x, N_y, N_z] = [512, 380, 380]$. A Reynolds number of 2800 is considered, where Re$= u_{\text{bulk}} h/\nu$. The bulk velocity u_{bulk} is given by the mean velocity at the hill crest, and h is the hill height. After reaching a statistically stationary point, the DNS data were averaged in the cross-stream (z–direction) and temporally for 44 flow through times.

A Linear Eddy Viscosity Model (LEVM) is frequently used to close the Reynolds stresses that appear in the RANS equations [17]. This closure takes the form

$$b_{ij} = -C_\mu \hat{S}_{ij}, \tag{3.17}$$

which will serve for comparison purposes as the 'existing' model.

As outlined in Section 3.2, the basis on which to train the model must first be identified. As previously derived [132], a minimal integrity basis for the anisotropy tensor can be formulated using the normalized mean rotation and shear rate tensors, $\hat{R}_{ij}$ and $\hat{S}_{ij}$, respectively. Since the anisotropy stress tensor is symmetric and deviatoric, each of $\mathcal{T}_{ij}^{(n)}$ must also have these properties. After formulating combinations of $\hat{S}_{ij}$ and $\hat{R}_{ij}$ with these properties, and owing to the Cayley-Hamilton theory, all symmetric and deviatoric tensors that are combinations of $\hat{S}_{ij}$ and $\hat{R}_{ij}$ can be formed as a linear combination of the 10 basis tensors shown in Table 3.3 [132].

$\mathcal{T}_{ij}^{(1)}$	$= \hat{S}_{ij}$	$\mathcal{T}_{ij}^{(6)}$	$= \hat{R}_{ik}\hat{R}_{kl}\hat{S}_{lj} + \hat{S}_{ik}\hat{R}_{kl}\hat{R}_{lj}$ $-\frac{2}{3}\hat{S}_{pk}\hat{R}_{kl}\hat{R}_{lp}\delta_{ij}$
$\mathcal{T}_{ij}^{(2)}$	$= \hat{S}_{ik}\hat{R}_{kj} - \hat{R}_{ik}\hat{S}_{kj}$		
$\mathcal{T}_{ij}^{(3)}$	$= \hat{S}_{ik}\hat{S}_{kj} - \frac{1}{3}\hat{S}_{lk}\hat{S}_{kl}\delta_{ij}$	$\mathcal{T}_{ij}^{(7)}$	$= \hat{R}_{ik}\hat{S}_{kl}\hat{R}_{lp}\hat{R}_{pj} - \hat{R}_{ik}\hat{R}_{kl}\hat{S}_{lp}\hat{R}_{pj}$
$\mathcal{T}_{ij}^{(4)}$	$= \hat{R}_{ik}\hat{R}_{kj} - \frac{1}{3}\hat{R}_{lk}\hat{R}_{kl}\delta_{ij}$	$\mathcal{T}_{ij}^{(8)}$	$= \hat{S}_{ik}\hat{R}_{kl}\hat{S}_{lp}\hat{S}_{pj} - \hat{S}_{ik}\hat{S}_{kl}\hat{R}_{lp}\hat{S}_{pj}$
$\mathcal{T}_{ij}^{(5)}$	$= \hat{R}_{ik}\hat{S}_{kl}\hat{S}_{lj} - \hat{S}_{ik}\hat{S}_{kl}\hat{R}_{lj}$	$\mathcal{T}_{ij}^{(9)}$	$= \hat{R}_{ik}\hat{R}_{kl}\hat{S}_{lp}\hat{S}_{pj} + \hat{S}_{ik}\hat{S}_{kl}\hat{R}_{lp}\hat{R}_{pj}$ $-\frac{2}{3}\hat{S}_{qk}\hat{S}_{kl}\hat{R}_{lp}\hat{R}_{pq}$
$\mathcal{T}_{ij}^{(10)} = \hat{R}_{ik}\hat{S}_{kl}\hat{S}_{lp}\hat{R}_{pq}\hat{R}_{qj} - \hat{R}_{ik}\hat{R}_{kl}\hat{S}_{lp}\hat{S}_{pq}\hat{R}_{qj}$			

Table 3.3: The ten tensor bases that exactly describe the anisotropic stress tensor.

Using this basis, the anisotropic stress tensor can be represented exactly as

$$b_{ij} = \sum_{n=1}^{10} G^{(n)} \mathcal{T}_{ij}^{(n)} \left(\hat{R}_{ij}, \hat{S}_{ij} \right). \tag{3.18}$$

In the case of statistically two-dimensional flows, as is the case here, the basis sim-

$\mathcal{T}_{ij}^{(1)}$	$\hat{S}_{ij}$
$\mathcal{T}_{ij}^{(2)}$	$\hat{S}_{ik}\hat{R}_{kj} - \hat{R}_{ij}\hat{S}_{kj}$
$\mathcal{T}_{ij}^{(3)}$	$\hat{S}_{ik}\hat{S}_{kj} - \frac{1}{3}\hat{S}_{lk}\hat{S}_{kl}\delta_{ij}$
λ_1	$\hat{S}_{lk}\hat{S}_{kl}$
λ_2	$\hat{R}_{lk}\hat{R}_{kl}$

Table 3.4: The reduced basis set for statistically two-dimensional flows.

plifies to only three tensors and the coefficients depend on at most only two invariants as shown in Table 3.4 [132, 136].

Following the sparse regression methodology described in Section 3.2, the DNS dataset is formulated into $\mathbb{D}$ and $\mathbb{T}$. However, instead of modeling b_{ij} directly, the anisotropic stress tensor is split into linear and nonlinear portions, denoted by $b_{ij}^{\parallel}$ and $b_{ij}^{\perp}$, respectively. The linear portion will be taken as the standard LEVM and the nonlinear portion will be the subject of modeling efforts.

$$b_{ij} = b_{ij}^{\perp} + b_{ij}^{\parallel} \tag{3.19}$$

$$= b_{ij}^{\perp}\left(k, \varepsilon, \hat{S}_{ij}, \hat{R}_{ij}\right) - C_{\mu}\hat{S}_{ij} \tag{3.20}$$

$$\mathcal{D}_{ij} = b_{ij}^{\perp} = b_{ij} + C_{\mu}\hat{S}_{ij}. \tag{3.21}$$

This strategy is employed based upon the recommendation of several works that have pointed out the ill-conditioning of the RANS equations [144]. These works suggest that separating the model into a linear portion (solved implicitly with the viscous terms in the RANS solver) and a nonlinear portion (solved explicitly) improves stability of the integrated RANS solver [30, 32]. Further, since the standard LEVM model is used as the starting point for modeling, the basis is formulated using data from a forward solution in OpenFOAM [145] using the LEVM closure. Because the $k - \varepsilon$ equations contain models and are thereby a source of error in the 'trusted' training data, this data must be used as a starting point for modeling.

Using this formulation, sparse regression is employed to discover an improved

model. This effort results in both an *a priori* and an *a posteriori* analysis of the model. In the former analysis, the training data is used to evaluate the accuracy of the learned model within the context of predicting the anisotropy tensor. In the latter analysis, the learned model is implemented in OpenFOAM and the forward solution is compared against the trusted DNS data and the existing LEVM. Additionally, as an 'upper end' metric, a look-up table was provided to the OpenFOAM RANS solver for the Reynolds stress terms that appear in both momentum and production in the $k-\varepsilon$ equations. This dataset serves as the performance of an ideal model that exactly captures the behavior of the Reynolds stresses while highlighting the model errors associated with the $k-\varepsilon$ model equations themselves.

Two learned models are discovered using sparse regression, one with three terms ($\lambda = 0$, denoted Learned 1) and the second with two terms ($\lambda = 15$, denoted Learned 2). Both learned models take the form,

$$b_{ij}^{\perp} = \left[\frac{1}{1000 + \lambda_1^3}\right] \left(C_1 \mathcal{T}^{(1)} + C_2 \mathcal{T}^{(2)} + C_3 \mathcal{T}^{(3)}\right), \tag{3.22}$$

and are detailed in Table 3.5.

The *a posteriori* analysis for the learned models includes an assessment of recirculation (Table 3.6) and velocity predictions for the training case (Figure 3.9) and a test case at a higher Reynolds number (Figure 3.11). These results are discussed in detail in Section 3.3.2.3.

3.3.2.2 A priori analysis

Each model developed can be assessed using the data with which it was trained. This represents an *a priori* assessment of the model, but has limitations as it does not take into account issues of stability or sensitivity that may be encountered within the context of a RANS solver. Further, since the forward solution is not computed

Model	C_1	C_2	C_3	$\epsilon^{\boldsymbol{b}}$	RMSE
LEVM	-	-	-	1.02	0.16
Learned 1	63.12	51.42	10.98	0.64	0.10
Learned 2	63.14	51.42	0	0.64	0.10

Table 3.5: Summary of learned model coefficients and *a priori* errors compared with the standard LEVM.

here, all assessments of model accuracy are computed with respect to the anisotropic stress tensor.

Shown in Figure 3.8, the standard LEVM does a reasonable job predicting the b_{12} component of the anisotropy tensor, but it struggles for the diagonal components. In all three cases, both the sign and magnitude are incorrect. The learned models, in contrast, capture the correct sign for the diagonal components and improve the magnitude inaccuracies present in the standard LEVM for the b_{12} component. However, for Learned 2, with the elimination of the third basis term, the prediction for b_{33} is also lost.

Using the L-2 norm as a metric for error, the learned models reduce model error in the anisotropic stress tensor by 41% with respect to LEVM.

3.3.2.3 A posteriori analysis

The true test of any model is its performance in the context of a forward solver. It is in this sense that model shortcomings become apparent, e.g., sensitivity or stability issues. Further, while the aim of Reynolds stress modeling is to improve accuracy in describing the stresses, the ultimate goal is that these models will improve predictions in the velocity field.

In order to assess the improvement of the learned model over the LEVM, the learned models were integrated into OpenFOAM, solved in conjunction with the $k-\varepsilon$ equations, and compared with the LEVM model and the $k-\varepsilon$ equations with a look-up table containing the DNS values for the anisotropy tensor. In each case, the RANS

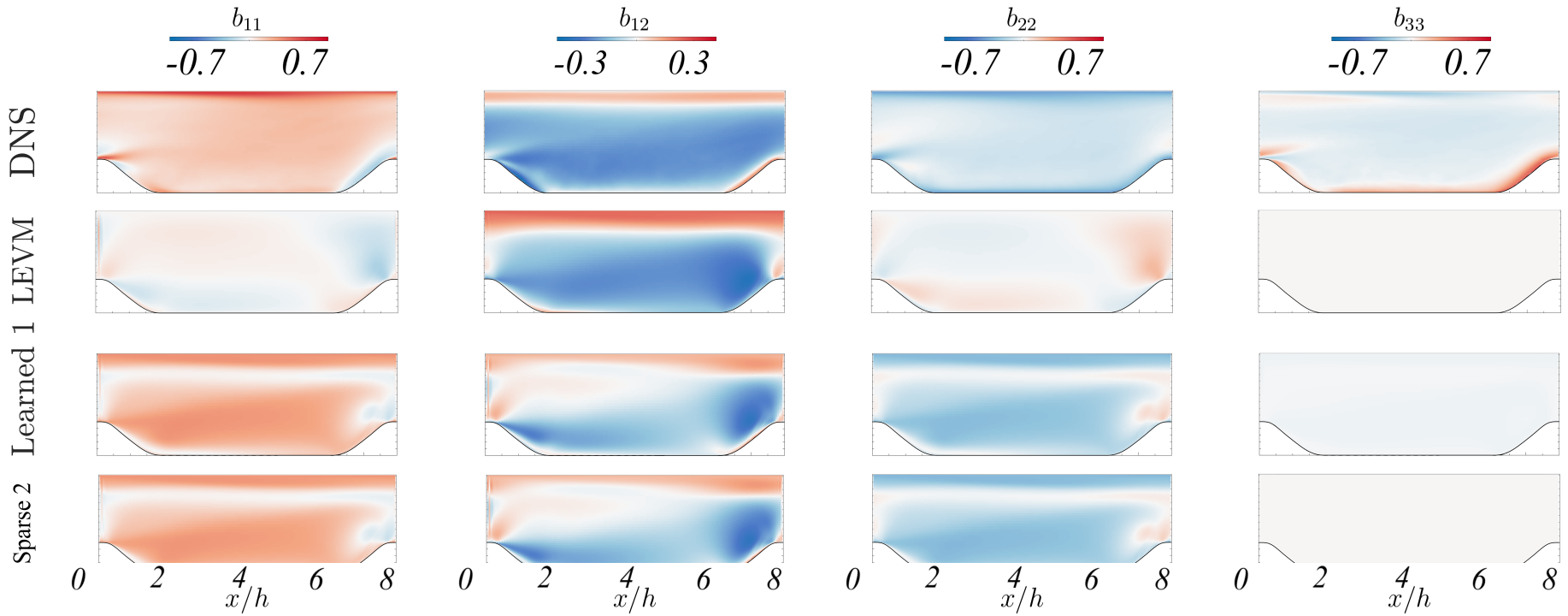
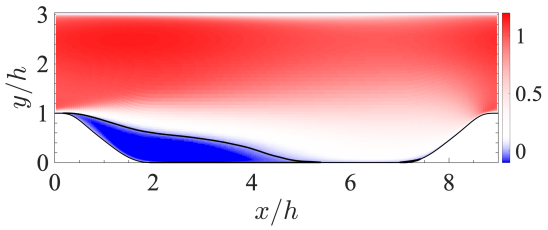
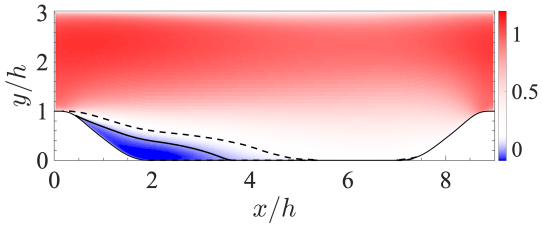


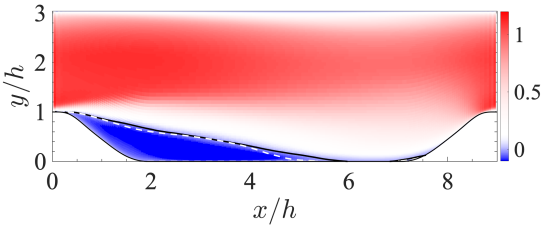
Figure 3.8: Comparisons of the components of the Reynolds stress tensor computed using DNS, LEVM, and the two learned models.



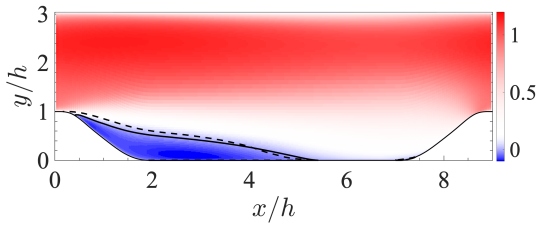
(a) DNS



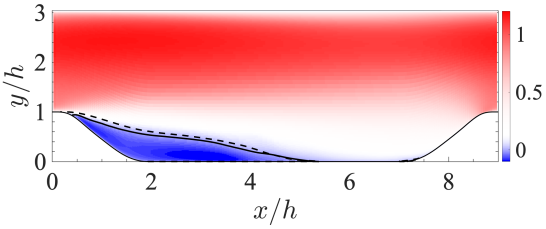
(b) LEVM



(c) Lookup table for b_{ij}



(d) Learned 1



(e) Learned 2

Figure 3.9: Forward solutions of the mean, normalized velocity, $\langle u \rangle / u_{\text{bulk}}$, for the standard LEVM model, the two learned models, the lookup table for DNS values of b_{ij} and the DNS results. The solid line represents the region of recirculation and the dashed line overlays where this region exists in the DNS data.

equations were solved on a two-dimensional grid of resolution $(N_x, N_y) = (200, 160)$ with the same physical dimensions as described in Breuer et al. [10] (and used for the DNS computations). Periodic conditions were imposed at the left and right faces, and 'patch' conditions were imposed on the front and back faces to enforce a 2-D solution. The bottom and top walls were treated as no-slip and a forcing term was added such that the velocity at the top of the hill crest enforced the desired Reynolds number.

The mean velocity normalized by the bulk velocity, u_{bulk}, is shown in Figure 3.9 and the detached regions are delineated by a black line. It is observed that LEVM under predicts recirculation compared with the DNS results (Figure 3.9(b)), while both

Reynolds number	Model	Primary		Secondary	
		Separation	Reattachment	Separation	Reattachment
Re= 2800	LEVM	0.43 (1.62)	3.64 (0.32)	–	–
	Learned 1	0.40 (1.53)	5.38 (0.005)	7.14 (0.04)	7.31 (0.008)
	Learned 2	0.40 (1.48)	5.38 (0.005)	7.14 (0.04)	7.31 (0.01)
	DNS	0.16 (–)	5.35 (–)	6.87 (–)	7.25 (–)
Re = 5600	LEVM	–	–	–	–
	Learned 1	0.40	5.38	7.14	7.30
	DNS [10]	0.18	5.14	–	–
	DNS [146]	0.17	5.04±0.09	7.04	7.31

Table 3.6: Summary of separation and reattachment locations for all models compared with DNS. Relative error with respect to the DNS values are shown in parentheses.

learned models demonstrate marked qualitative improvement in velocity prediction. Quantitative measurements of separation and reattachment locations for both the primary and secondary recirculation regions are detailed in Table 3.6. The learned models predict both primary and secondary reattachment points within 1% of the DNS values, with exception of the primary separation point. In comparison, LEVM under predicts the primary reattachment point by 32% compared with DNS and fails to predict existence of the secondary recirculation.

Examination of the momentum RANS equation (Eq. (3.4)) makes clear that $\langle u'u' \rangle$ and $\langle u'v' \rangle$ are the only Reynolds stress components that contribute to $\langle u \rangle$ and therefore to the prediction of recirculation. By examining b_{11} and b_{12} in Figure 3.10, it can be seen that both components of anisotropy contribute to the prediction of the separation location, however the b_{12} component is most important for the prediction of reattachment. This can be seen by considering the areas of high gradients (since the contribution to the velocity field is in the form of the divergence of the Reynolds stresses), specifically the stream-wise gradient for the 11-component and the vertical gradient for the 12-component of the model are important. As shown in Figure 3.10, the second basis tensor, $\mathcal{T}_{ij}^{(2)}$, is the most important contribution for accurately describing b_{11} and b_{22} and the first basis tensor, $\mathcal{T}_{ij}^{(1)}$, is the most dominant contribution for modeling b_{12}. Thus $\partial \mathcal{T}_{11}^{2}/\partial x$ and $\partial \mathcal{T}_{12}^{(1)}/\partial y$ are the most important for accurately predicting the recirculation region. Finally, the third basis is critical for accurately describing the b_{33} component, though for this particular configuration (since z is a homogeneous direction), accuracy in this component is not required for predicting the statistically two-dimensional mean flow field.

3.3.2.4 Application outside the scope of training

To assess the range of application of the learned model (Learned 1), a forward RANS simulation was conducted in OpenFOAM for two configurations and Reynolds

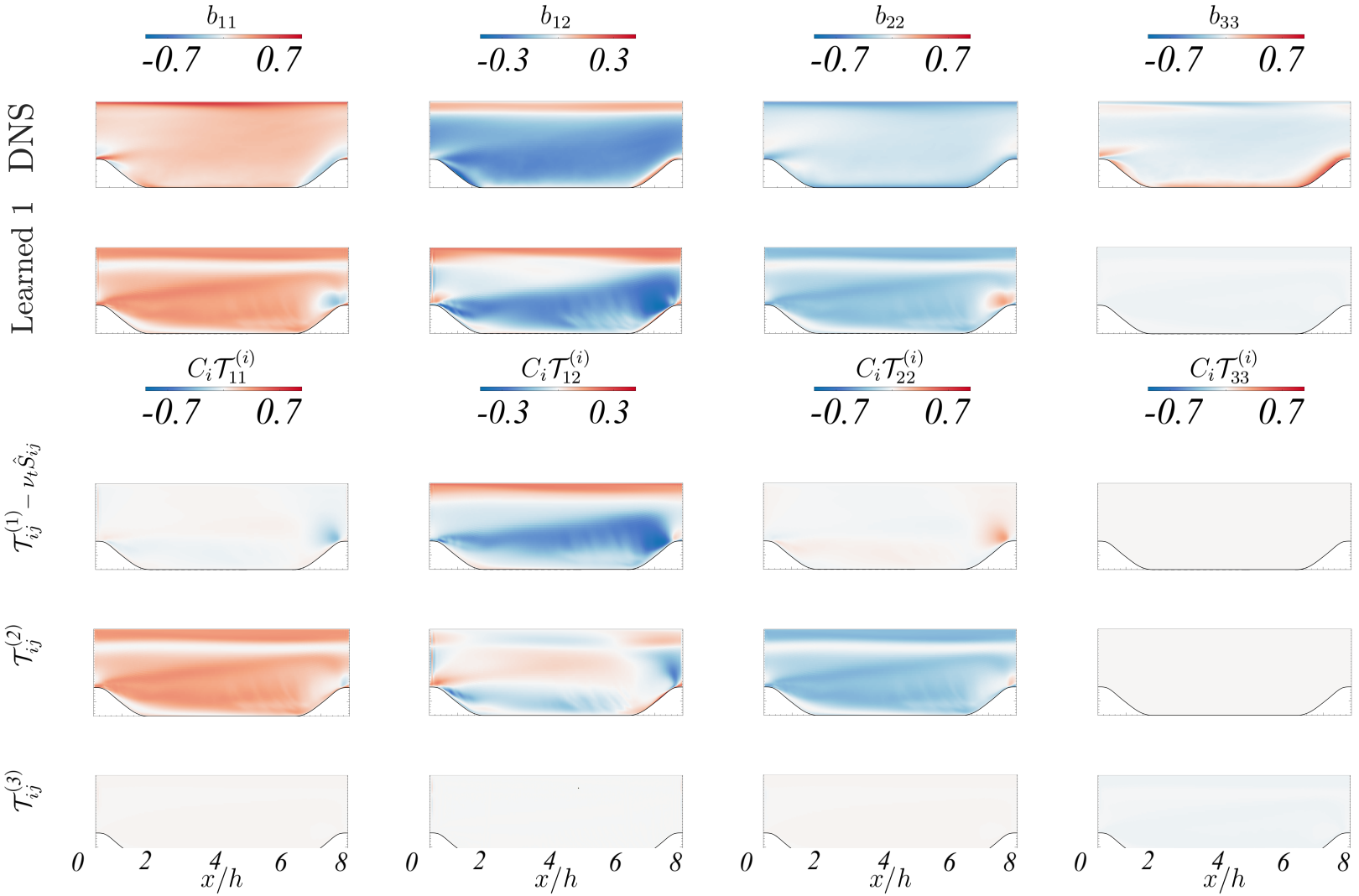


Figure 3.10: Contributions to each component of the anisotropy tensor from each of the bases for the model Learned 1, compared with the DNS field

numbers outside the scope of its training: (1) the wavy wall configuration but at a higher Reynolds number and (2) flow over a backward facing step, where massive separation is also observed. In both of these configurations, DNS and/or experimental data is available to assess model performance. As in the previous section, the learned model's performance is compared against the LEVM.

Since full-field data is not available for these additional cases, model performance is determined based on prediction of the reattachment point in the flow. These results are summarized in Table 3.7.

The first out-of-scope configuration considered is the periodically constricted channel configuration as described in the previous section, but at a higher Reynolds number. More exhaustive details on the Reynolds number dependencies for this configuration can be found in Breuer et al. [10], but to briefly summarize, increasing the Reynolds number for this configuration results in differences in recirculation size as well as in separation and reattachment locations. The selection of Re= 5600 was chosen due to the existence of available DNS data. Here, the learned model (Learned 1) was again implemented in OpenFOAM and compared against the openly available dataset provided by Breuer et al. [10], in which only a primary recirculation region is observed. Improvement over LEVM is observed in this case, as shown in Figure 3.11 (a) and Table 3.7. The LEVM solution again under predicts recirculation, while the learned model predicts the reattachment location within 5% of the DNS value reported in Breuer et al. [10] and within 5-9% of the DNS value reported in Krank et al. [146]. The primary separation location is marginally over two times further in the stream-wise direction as compared to both DNS results [10, 146]. The learned model also predicts the small secondary recirculation region that is reported in Krank et al. [146]. Breuer et al. [10] does not observe this secondary recirculation, however this appears to be due to differences in numerical schemes and order of accuracy as compared with Krank et al. [146]. In this secondary region, the learned model pre-

dicts the separation and reattachment points within 1% and 0.1%, respectively, as compared with the DNS reported in Krank et al. [146].

The second configuration considered to assess model performance outside the scope of its training is turbulent flow over a backward-facing step for Re= 5000. For this configuration, model performance is assessed by comparison with reported DNS values [11] and experimental values [12, 147] as shown in Figure 3.11 and quantified in Table 3.7. The same configuration is used as described in these works, and the Reynolds number is defined using the step height ($h = 9.6$ mm). No-slip boundary conditions were enforced at the top and bottom walls in the RANS simulations, and a fixed velocity condition is enforced at the inlet and a zero gradient condition for the outflow. 'Patch' conditions are implemented on the front and back surfaces to enforce a 2-D flow.

LEVM is found to under predict the reattachment point between 23 and 32%, while the model trained on the periodic hill data in previous sections predicts reattachment within 5% as compared to the DNS and experimental values.

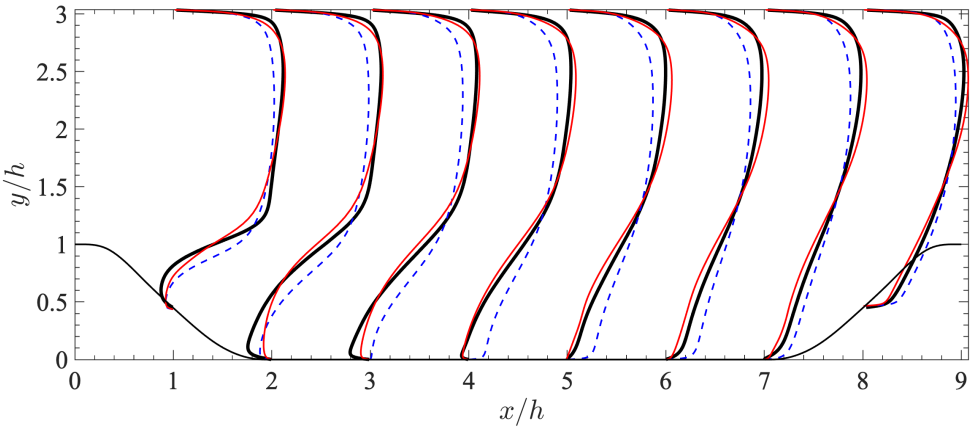
It is notable that the level of performance of the learned model outside the scope of it's training, particularly in the case of the backward-facing step, is comparable to the out-of-scope performance of the Tensor Basis Neural Network developed by Ling et al. [28]. It is also relevant to point out that Ling et al. [28] trained on six cases to achieve this level of performance, while the present model was trained only on flow through a periodically constricted channel at Re= 2800.

3.3.2.5 Modeling with sparse data

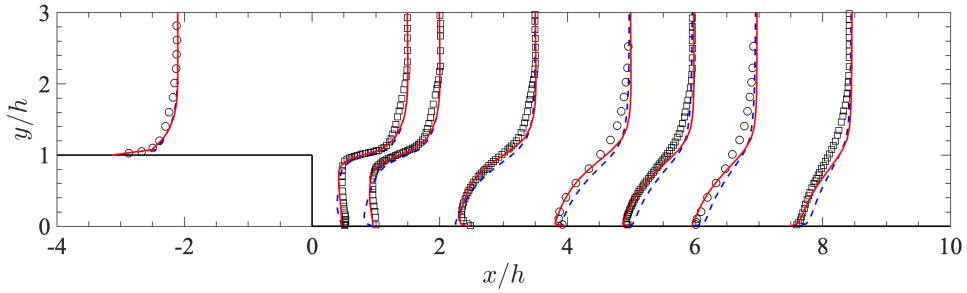
Because the sparse regression methodology seeks to uncover underlying physics, far less data is required to achieve reasonable learned models. This is demonstrated in two contexts. First, a model is learned using only the $y-$dependent data along eight stream-wise locations (see Table 3.8). Due to the grid spacing of the RANS simulation,

Configuration	Re	Model	Primary		Secondary	
			Separation	Reattachment	Separation	Reattachment
Periodic hills	2800	LEVM	0.43 (1.62)	3.64 (0.32)	–	–
		Learned 1	0.40 (1.53)	5.38 (0.005)	7.14 (0.04)	7.31 (0.008)
		Learned 2	0.40 (1.48)	5.38 (0.005)	7.14 (0.04)	7.31 (0.01)
		Learned 3	0.34 (1.14)	5.54 (0.04)	–	–
		Learned 4	0.71 (3.43)	4.98 (0.07)	7.15 (0.04)	7.50 (0.03)
		DNS	0.16	5.35	6.87	7.25
	5600	LEVM	0.40 (1.35)	3.01 (0.40)	–	–
		Learned 1	0.40 (1.38)	5.38 (0.05)	7.14 (0.01)	7.30 (0.001)
		Learned 3	0.34 (0.98)	5.41 (0.07)	–	–
		Learned 4	0.41 (1.41)	4.55 (0.09)	–	–
		DNS [10]	0.18	5.41	–	–
		DNS [146]	0.17	5.04±0.09	7.04	7.31
		Experiment [13]	–	4.83	–	–
	10600	LEVM	0.41 (1.30)	3.78 (0.26)	–	–
		Learned 1	0.40 (1.24)	5.10 (0.002)	–	–
		Learned 3	0.33 (0.85)	5.59 (0.10)	–	–
		Learned 4	0.40 (1.24)	4.60 (0.10)	–	–
		LES [10]	0.19	5.09	–	–
		Experiment [13]	–	4.21	–	–
Backward-facing step	5000	LEVM	0.49	4.52 (0.28)	–	–
		Learned 1	0.90	6.17 (0.02)	–	–
		DNS [11]	–	6.28	–	–
		Experiment [12]	–	6 ± 0.15	–	–
		Experiment [147]		6.51	–	–

Table 3.7: Summary of separation and reattachment locations for all models compared with DNS. Relative error with respect to the DNS values are shown in parentheses. Dashes indicate the data is either not reported or not observed. Note that models 'Learned 3' and 'Learned 4' are discussed in Section 3.4.



(a) Flow through periodically constricted channel at Re= 5600



(b) Flow over a backward-facing step at Re= 5000

Figure 3.11: Velocity profiles for flow through a periodically constricted channel at Re= 5600 (a) and flow over a backward-facing step at Re= 5000 (b). Learned 1 (—), LEVM (- - -), DNS of [10] (▬), DNS of [11] (□) and experiments of [12] (○)

only 160 data points were used for training of each case (compared with 32,000 when using the full dataset in the previous section). As seen in Table 3.8, similar model performance is observed for models trained using data located at $x/h = (4-8)$ when compared with the model learned using the full dataset. Secondary recirculation is predicted in three of these training sets. Interestingly, the model trained at $x/h = 8$, where recirculation is not present, is able to predict recirculation in both regions of flow separation. Additionally, it is notable that the model is insensitive to variation in coefficients, especially for the first and third terms.

To further assess the performance of sparse regression in using sparse data, subsets of data are randomly chosen throughout the domain and used as training data. Datasets ranging from 30,000 to 50 training points were assessed (see Table 3.9).

Training	Learned coefficients			Error	Primary		Secondary	
(x/h)	C_1	C_2	C_3	$\epsilon^{\langle u \rangle}$	separation	reattachment	separation	reattachment
1	32.35	45.42	19.69	0.17	0.36 (1.24)	4.15 (0.22)	–	–
2	36.72	46.85	36.43	0.16	0.35 (1.23)	4.27 (0.20)	–	–
3	51.81	48.75	37.59	0.15	0.34 (1.17)	4.78 (0.11)	–	–
4	51.38	53.05	40.99	0.12	0.40 (1.52)	5.32 (0.01)	7.22 (0.05)	7.35 (0.01)
5	48.63	55.83	35.85	0.13	0.41 (1.56)	5.10 (0.04)	7.17 (0.04)	7.18 (0.01)
6	49.85	55.77	23.34	0.13	0.41 (1.56)	5.11 (0.04)	–	–
7	58.34	54.19	-4.04	0.13	0.41 (1.56)	5.14 (0.04)	–	–
8	112.56	51.42	-50.00	0.12	0.37 (1.31)	4.35 (0.18)	7.10 (0.03)	7.27 (0.002)
			Learned 1	0.12	0.40 (1.53)	5.38 (0.005)	7.14 (0.04)	7.31 (0.008)
			LEVM	0.17	0.43 (1.61)	3.64 (0.32)	–	–
			DNS	–	0.16 (–)	5.35 (–)	6.87 (–)	7.25 (–)

Table 3.8: Summary of learned coefficients using sparse data, i.e. only the y-dependent data at the specified x/h location. Model error is reported for the *a posteriori* velocity, and separation and reattachment points are compared with DNS, LEVM, and the 'Learned 1' model for Re= 2800.

	Coefficients			
n^{train}	C_1	C_2	C_3	$\epsilon^{\boldsymbol{b}}$
$nx \times ny$	63.12	51.42	10.98	0.64
30,000	62.50	51.52	11.24	0.64
20,000	52.50	45.77	12.19	0.65
10,000	42.03	38.84	16.92	0.67
5,000	37.35	36.37	19.69	0.69
1,000	33.32	35.41	21.76	0.69
500	33.91	35.07	20.33	0.69
100	33.00	32.5	23.34	0.71
50	31.14	38.28	20.66	0.68

Table 3.9: Summary of the learned coefficients for sparse, randomly sampled data using n^{train} training points. The error reported is the *a priori* error in the anisotropic stress tensor.

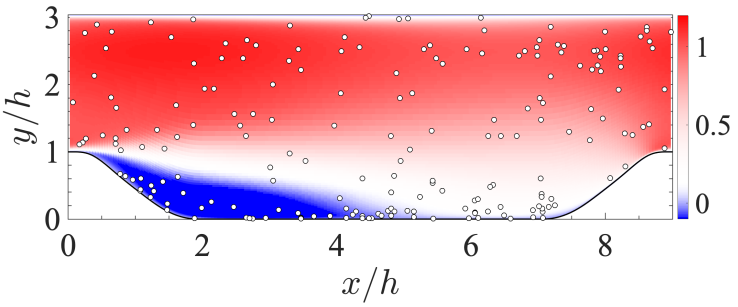


Figure 3.12: Example of the random points used for training set corresponding to $n^{train} = 200$. Location of points used for training ($\circ$), $\langle u \rangle$ from DNS (color).

While the learned coefficients change as the dataset is reduced, the *a priori* model error in the anisotropic stress tensor only increases by 8%. This suggests that sparse regression would make an excellent modeling construct for extremely sparse datasets, such as those available from experiments where obtaining a high level of resolution is challenging.

3.4 Training the model with experimental data

For many practical systems, procuring highly resolved computational data (i.e., DNS or highly resolved LES) is not feasible. Thus, in these cases modeling efforts are directed toward experimental data, which is inherently both sparse and noisy. To this end, we demonstrate the ability of the sparse regression methodology to successfully model both sparse and noisy data by using the Particle Image Velocimetry (PIV) data available from the Rapp and Manhart [13] experiments for turbulent flow through a periodically constricted channel. This work is the experimental analogy to Breuer et al. [10] and uses the same configuration described in Section 3.3.2.

We consider two cases (Re= 5600 and 10600) for which highly resolved computational data (either DNS or LES) are available [10] in addition to the experimental data [13]. The reason for this is two-fold. First, because the PIV measurements do not report $\langle w'w' \rangle$ or k, either of which is required for determining the anisotropy tensor, we rely on using the k value from the computational data to estimate $\langle w'w' \rangle$ and complete the experimental dataset. More exactly, we employ the relation, $\langle w'w' \rangle = 2k - \langle u'u' \rangle - \langle v'v' \rangle$, where k is taken from the computational datasets and $\langle u'u' \rangle$ and $\langle v'v' \rangle$ are supplied by the PIV measurements. This allows for the noise of the PIV measurements to be incorporated into the estimate for $\langle w'w' \rangle$.

In addition to completing the experimental dataset, the computational data also allows for the more systematic isolation of the effect of noise on modeling. To this end, we first train the model using the computational data (which is reported at 10 stream-wise locations), interpolated to the same physical locations as the experimental data to ensure equivalent sparsity in the dataset.

Following a similar procedure as was performed in Section 3.3.2.5, we find that a subset of the data (taken from all 10 stream-wise locations and vertical locations corresponding to $y/h \geq 2$ and resulting in a dataset of 591 points for Re= 5600 and 593 for Re= 10600) results in the most accurate model with respect to the *a priori*

Model name	Training set	Learned Coefficients C_1	C_2	C_3
Learned 1	DNS (full field), Re= 2800	63.12	51.42	10.98
Learned 3	DNS/LES (sparse), Re= 5600, 10600 [10]	61.72	44.07	12.58
Learned 4	Experiment, Re= 5600, 10600 [13]	58.56	55.51	118.89

Table 3.10: Summary of the learned model coefficients trained using the three training datasets: full-field DNS at Re= 2800, sparse DNS/LES data at Re= 5600 and 10600, and experimental data.

L-2 training error in b_{ij}. This model is termed 'Learned 3' and shown in Table 3.10. The coefficients are within 15% of the 'Learned 1' model, despite using a training set containing only 4% of the points as used in the full-field DNS training. The resulting recirculation predictions for Re= $(2800, 5600, 10600)$ are shown in Table 3.7 and compared against the LEVM and Learned 1 models as well as the available computational and experimental results from the literature.

Next, the experimental dataset [13] is used for training (using the same reduced sample locations as in training with the DNS/LES dataset). This results in a model with differing coefficients, especially in the case of C_3, though as reported previously, we observe that the resulting flow field is the least sensitive to $\mathcal{T}^{(3)}$ and thus this difference does not have large implications for the model's ability to predict recirculation (see Table 3.7). The resultant mean velocity profiles for all three learned models describe in Table 3.10 are compared against LEVM as well as the experimental and DNS values in Figure 3.13. It is evident from these plots that the learned models improve prediction of the reattachment location over LEVM as well as the free flow in the remainder of the domain. In summary, the sparse regression methodology is capable of providing algebraic closure of the terms appearing in the RANS equations with marked improvement over existing models, even when trained on sparse experimental measurements.

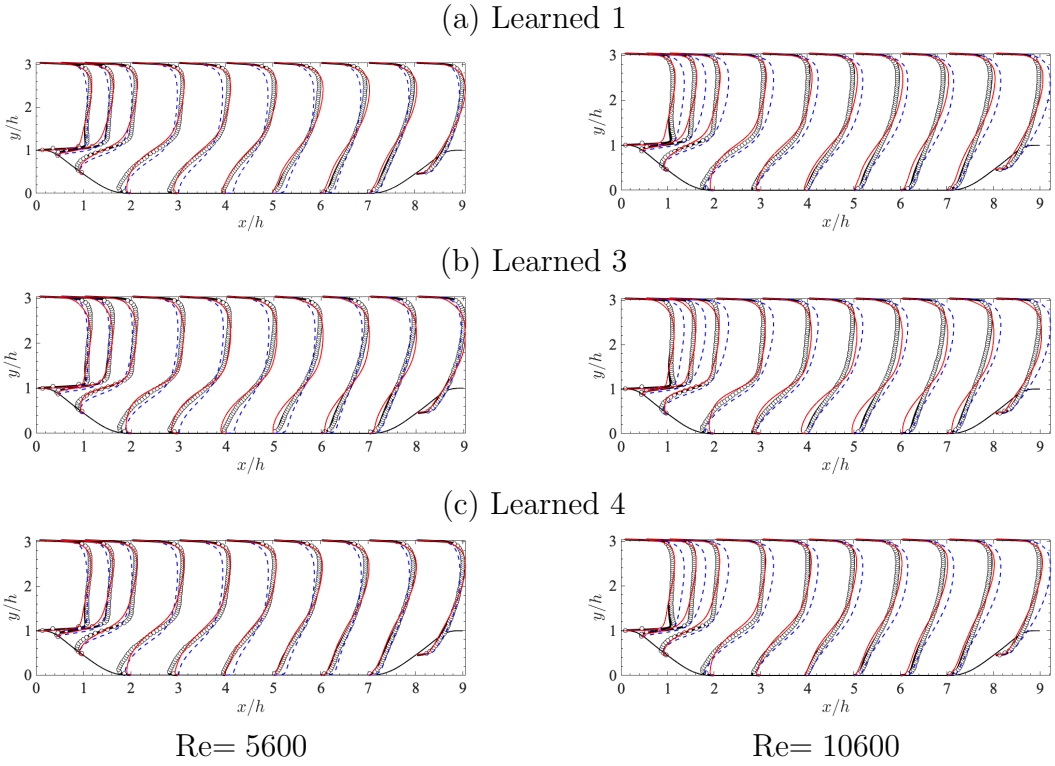


Figure 3.13: Velocity profiles for flow through a periodically constricted channel at Re= 5600 (left column) and Re= 10600 (right column). Each plot shows the learned model denoted in the caption (—), LEVM (- - -), DNS/LES of [10] (▬), experimental values from [13] (○).

3.5 Conclusion

In this work, a turbulence closure modeling methodology has been proposed as an alternative to other machine learning techniques, such as NNs. This method is based upon sparse regression which uses an L-2 norm with an L-1 norm penalty cost functional to produce a compact, algebraic model. Further, the inputs to the optimization algorithm are specifically tailored in order to ensure form invariance. This is specifically accomplished by arranging the trusted and basis tensorial data into column vectors, thereby constraining coefficients to be invariant with respect to direction. By generating a model in this form, several important modeling properties can be achieved: form (or Galilean) invariance, interpretability, and ease of dissemination. Using two canonical cases, it was demonstrated that this technique produces results with model accuracies similar to that of modern NN methodologies, even when using a drastically reduced training dataset.

Using homogeneous free shear turbulence as a preliminary example, sparse regression was able to return the LRR-IP model used to generate a synthetic dataset, even when large amounts of noise were applied. Next, using DNS data for homogeneous free shear turbulence, sparse regression learned a model that reduces model error by 70% as compared to the existing LRR-IP and LRR-QI models.

In the case of turbulent flow through a periodically constricted channel, sparse regression uncovered a model that has comparable performance to a modern NN considering the same flow, however this performance can be achieved using a drastically minimal dataset and the resultant model form is available in a compact, algebraic form. Additionally, the learned model demonstrated significant improvements in performance as compared with LEVM for a much higher Reynolds number, and outside the scope of its training. Further, due to the ability of sparse regression to learn predictive models using minimal datasets and noisy data (as demonstrated in Section 3.3.1), it is an ideal candidate for translating experimental data, which may be

both noisy and sparse, into accurate models.

Finally, sparse regression assumes complete generality and thus does not strictly require an existing model upon which to augment. This is an important property for other open areas of research, e.g., modeling multiphase turbulence [39, 40, 41, 127, 6], for which existing models are either unavailable or too inaccurate to reliably use as a baseline model upon which to build. Such an approach can also be applied to turbulent combustion, in which heat release due to chemical reactions give rise to 'back scatter' and existing models based on an energy cascade fail to be predictive [43, 44].

In future work, integration of Gene Expression Programming (GEP) with sparse regression may be beneficial for cases in which complex algebraic dependencies upon the principal invariants become necessary (i.e., coefficients that are either constant or have simple dependencies on the invariants do not reduce model error sufficiently). In this event, sparse regression would be employed to determine the most important basis tensors, and then GEP could be used to determine the functional dependence of coefficients on the principal invariants.

CHAPTER IV

Application to Gas-Solid Flows

In this work, model closures of the multiphase Reynolds-Average Navier–Stokes (RANS) equations are developed for homogeneous, fully-developed gas–particle flows. To date, the majority of RANS closures are based on extensions of single-phase turbulence models, which fail to capture complex two-phase flow dynamics across dilute and dense regimes, especially when two-way coupling between the phases is important. In the present study, particles settle under gravity in an unbounded viscous fluid. At sufficient mass loadings, interphase momentum exchange between the phases results in the spontaneous generation of particle clusters that sustain velocity fluctuations in the fluid. Data generated from Eulerian–Lagrangian simulations are used in a sparse regression method for model closure that ensures form invariance. Particular attention is paid to modelling the unclosed terms unique to the multiphase RANS equations (drag production, drag exchange, pressure strain and viscous dissipation). A minimal set of tensors is presented that serve as the basis for modelling. It is found that sparse regression identifies compact, algebraic models that are accurate across flow conditions and robust to sparse training data.

4.1 Introduction

Many natural and industrial processes involve the flow of solid particles, liquid droplets or gaseous bubbles whose dynamical evolution and morphology are intimately coupled with a carrier fluid. A peculiar behaviour of disperse multiphase flows is their ability to give rise to large-scale structures (hundreds to thousands of times the size of individual particles), from dense clusters to nearly particle-free voids (see figure 4.1). The emergence of spatial segregation in particles can be attributed to a number of factors, e.g. due to dissipation during inelastic collisions [148, 149], viscous damping by the fluid [150], preferential concentration of particles by the background turbulence [69] and instabilities that arise due to interphase coupling [68, 67, 40, 41]. Such large-scale heterogeneity can effectively 'demix' the underlying flow, reducing contact between the phases resulting in enormous consequences at scales much larger than the size of individual particles, such as hindered heat and mass transfer [70, 151, 152, 56, 6].

The focus of the present work is on modelling disperse multiphase turbulence at length scales much larger than the particle diameter. Such flows can be categorized into two broad classes: (i) flows where the turbulence originates in the continuous phase with the discrete phase modulating the small scales of the turbulence [see, e.g., 66, and references therein]; and (ii) flows where the turbulence arises due to the coupling between the discrete and continuous phases. The former is mainly focused on how the discrete phase modifies the classical turbulence structures seen in single-phase flows. The latter is the focus of the present work and can be observed in gas–particle flows when the mass of the particles is of the same order or greater than that of the gas phase, or in bubbly flows when the bubble volume fraction is high enough to lead to buoyancy-driven instabilities [e.g. 153, 154, 155].

To date, turbulent particle-laden flows are most often discussed in the dilute limit where the fluid-phase turbulence interacts with inertial particles without significant

feedback from the particles. It is well established that dilute suspensions of heavy particles in isotropic turbulence will preferentially concentrate in regions of high strain rate and low vorticity [69]. In the presence of a mean body force (e.g. gravity), particles have been observed to experience enhanced settling as a result of preferential sweeping, by which the particles tend toward regions of downward fluid motion when encountering vortical structures in the flow [e.g. 156, 157, 158, 159, 160].

When the particle concentration is sufficiently high, the background flow is largely controlled by interphase coupling. Under these conditions, particles tend to accumulate in regions of low vorticity, resembling preferential concentration that typically occurs in dilute particle-laden turbulence. However, in the dense limit the vorticity is generated in shear layers between highly concentration regions (clusters), unlike in classical preferential concentration where vorticity would exist even in the absence of the disperse phase [41]. Additional effects contribute to the settling velocity and spatial segregation of particles in denser suspensions when two-way coupling between the phases is non-negligible.

Seminal works by G. K. Batchelor have provided theoretical estimates describing the motion of collections of solid particles suspended in viscous flows [161, 162], in addition to important insights on the instabilities present in such systems. For example, Batchelor [163] demonstrated that small rigid spheres falling under gravity will give rise to long-range hydrodynamic interactions that result in hindered settling [161]. In more recent studies, it was demonstrated that at higher Reynolds numbers and particle concentrations, momentum exchange between the phases results in *enhanced* settling when the mean mass loading, φ, defined by the ratio of the specific masses of the particle and fluid phases, is of order one or larger [41]. In statistically homogeneous gravity-driven gas–solid flows, the average particle settling speed, $\mathcal{V}$, can be approximated as

$$\mathcal{V} = \mathcal{V}_0 + \langle u_f \rangle_p \tag{4.1}$$

for Stokes flow [41], where $\mathcal{V}_0 = \tau_p g$ is the terminal Stokes settling velocity of an isolated particle with τ_p the particle response time and g gravity. In this expression, the phase-averaged fluid velocity, $\langle u_f \rangle_p = \langle \varepsilon_p u_f \rangle / \langle \varepsilon_p \rangle$, is sometimes referred to as the velocity *seen* by the particles, where u_f is the local fluid velocity aligned with gravity, ε_p is the local particle volume fraction, and angled brackets denote a spatial and temporal average. At sufficient mass loading, fluctuations in particle concentration can generate and sustain fluid-phase turbulence (as shown in figure 4.1), referred to here as cluster-induced turbulence (CIT). Because clusters entrain the carrier phase, u_f and ε_p are often highly correlated, resulting in $\mathcal{V} > \mathcal{V}_0$.

Due to the breadth of length- and time-scales present in turbulent fluid–particle mixtures, accurate modelling of industrial and environmental flows remains challenging. Thus, the Reynolds-averaged Navier–Stokes (RANS) equations are the workhorse of industry to inform engineering designs and decisions. Because of the importance of the multiphase physics present in large-scale systems, developing multiphase RANS closures that are accurate under relevant conditions is critically important.

To date, multiphase turbulence models have largely relied upon extensions to single-phase models (e.g. [45, 46, 47, 48, 49, 50, 51, 52, 53]) that were derived directly from the Navier–Stokes equations. It should be noted, however, that multiphase turbulence does share some commonalities with single-phase flows, especially with variable-density turbulence. For example, multiphase flows subject to mean shear can develop velocity fluctuations that strongly modify the mean velocity profiles and transport properties of the flow [164]. Single-phase, variable-density flows subject to Rayleigh–Taylor instabilities [see, e.g. 165, 166] develop velocity and density fluctuations similar to those observed in heterogeneous bubbly flows [167]. However, the main difference between disperse multiphase and variable-density flows is that the former has a separate velocity for each phase while the latter has a single fluid velocity. Moreover, momentum coupling by particles introduces velocity fluctuations at small

scales that further complicates the energy budget of turbulence. By introducing the slip velocity, i.e., the velocity difference between the discrete and continuous phases, additional dimensionless parameters, such as the Stokes number and mass loading, are needed to describe multiphase turbulence.

In contrast to modelling by analogy with single-phase flow, Fox [39] developed the exact Reynolds-averaged equations for collisional fluid–particle flows. In that work, it was demonstrated that directly averaging the Navier–Stokes equations fails to capture important two-phase interactions. Instead, it was demonstrated that phase averaging the mesoscale (locally averaged) equations results in a set of equations that explicitly account for two-way coupling contributions. Capecelatro et al. [41] further developed the Reynolds-averaged formulation of Fox [39] to include transport equations for the volume-fraction variance, drift velocity and the separate components of the Reynolds stresses of each phase and particle-phase pressure tensor. While exact, it does lead to a large number of unclosed terms that require modelling, which is the focus of the present study.

Accurate modelling of the unclosed terms that remain predictive from dilute to dense regimes remains an outstanding challenge. Fox [39] proposed closures of the phase-averaged (PA) terms based largely on single-phase turbulence models without extensive validation. [42] extended these models to account for near-wall effects in particle-laden channel flows. Agreement with the turbulence statistics obtained from simulation data was found to be satisfactory at first order (e.g. PA velocities) but less so at second order (e.g. PA turbulent kinetic energy). Innocenti et al. [55] drew upon a probability-density-function approach, along with extensions from single-phase turbulence modelling (particularly in the fluid phase), showing satisfactory agreement for statistics up to second order. However, the model was restricted to relatively dilute flows. Due to the large parameter space associated with turbulent multiphase flows, a reliable modelling approach valid across two-phase flow regimes (e.g. dilute

to dense limit) remains elusive.

Broadly speaking, extracting new models and understanding of physics from data has a long history in many diverse areas of science and engineering [see e.g. 168]. In the last decade, these data-driven techniques have been applied to turbulence modelling in several ways, including uncertainty prediction and quantification, model calibration and augmentation and the generation of entirely new models. Several recent works have utilized machine learning (neural networks are particularly popular [22, 23, 24, 25, 26, 169, 27, 28, 29, 30]) in order to translate large amounts of experimental or computational data into model closures. Neural networks have shown relatively exceptional performance outside the region in which they were trained. As a departure from more traditional modelling techniques, these methods are inserted modularly, as a 'black box,' into an existing flow solver. Thus, while they have displayed a high level of performance on a wide range of flow conditions, the closure does not satisfy the interpretability condition necessary for making physical inferences. Further, a large number of neural network approaches attempt to augment or correct existing models. However, as discussed above, in the context of multiphase flows appropriate existing models in which to augment do not exist.

Rather than relying on a best-fit strategy, as done in neural networks, Brunton et al. [38] developed a strategy based on sparse regression that identifies the underlying functional form of the nonlinear physics by optimizing a coefficient matrix that acts upon a matrix of trial functions. While this method requires knowledge about the physics of the system under configuration (in order to make informed selections of the trial functions), it can be reasonably assumed that the modeller is not entirely naive. In fact, traditional modelling techniques have relied nearly exclusively on this notion. Schmelzer et al. [170], Beetham and Capecelatro [58] recently extended the sparse identification framework of Brunton et al. [38] to infer algebraic stress models for the closure of RANS equations. In [170] the models are written as tensor polynomials

and built from a library of candidate functions. In [58] Galilean invariance of the resulting models are guaranteed through thoughtful tailoring of the feature space.

In this work, the sparse identification modelling framework of Beetham and Capecelatro [58] is employed to develop multiphase closure models for homogeneous, gravity-driven gas–solid flows. Eulerian–Lagrangian simulations are performed across a range of Archimedes numbers and volume fractions to provide training data. The terms appearing in the multiphase RANS equations recently derived in Capecelatro et al. [41] are extracted. We then build a minimally invariant basis set of tensors (i.e. a set of functional groups that serve as candidate terms in the desired model). Such basis sets are well established for single-phase turbulent flows [133, 136, 132]; however, an analogous basis has not yet been determined for multiphase flows. Using this basis and the sparse regression methodology, the compact functional form of the physics-based closures are inferred. As we consider exclusively statistically stationary and homogeneous systems, model realizability [17] is left for future work.

4.2 System description

4.2.1 Configuration under study

In the present study, rigid spherical particles of diameter d_p and density ρ_p are suspended in an unbounded (triply periodic) domain containing an initially quiescent gas of density ρ_f and viscosity ν_f. Gravity g acts in the negative x-direction. As particles settle, they spontaneously form clusters. Due to two-way coupling between phases, particles entrain the fluid, generating turbulence therein. A frame of reference with the fluid phase is considered, such that the mean streamwise fluid velocity is null. Given the relative simplicity of the configuration, only a few non-dimensional groups arise. An important non-dimensional number is the Archimedes number, defined as

$$\mathrm{Ar} = (\rho_p/\rho_f - 1) d_p^3 g/\nu_f^2. \tag{4.2}$$

	Dimensional Quantities		
$\mathcal{V}_0$	Stokes settling velocity	[m/s]	(0.02, 0.06, 0.2)
$\mathcal{L}$	Characteristic cluster length	[m]	$(5.0, 15.0, 50.0) \times 10^{-4}$
τ_p	Drag time	[s]	0.025
ρ_p	Particle density	[kg/m^3]	1000
d_p	Particle diameter	[m]	90×10^{-6}
ρ_f	Fluid density	[kg/m^3]	1
ν_f	Fluid viscosity	[m^2/s]	1.8×10^{-5}
g	Gravity	[m/s^2]	(0.8, 2.4, 8.0)
	Non-dimensional Quantities		
N_p	Number of particles		(610,370, 15,564,442, 30,518,514)
$\langle \varepsilon_p \rangle$	Mean volume fraction $\times 10^{-2}$		(0.1, 2.55, 5.0)
φ	Mean mass loading		(1.0, 26.2, 52.6)
Fr	Froude number		(5.6, 16.7, 55.6)
Ar	Archimedes number		(1.8, 5.4, 18.0)
	Computational Quantities		
	Domain size	[m]	$0.158 \times 0.038 \times 0.038$
	Grid size		$512 \times 128 \times 128$
	$L_x/\mathcal{L}$		316,105,32

Table 4.1: Summary of parameters for the configurations under consideration.

Alternatively, a Froude number can be introduced to characterize the balance between gravitational and inertial forces, defined as $\mathrm{Fr} = \tau_p^2 g/d_p$, where $\tau_p = \rho_p d_p^2/(18\rho_f \nu_f)$ is the particle response time. The Stokes settling velocity for an isolated particle is given by $\mathcal{V}_0 = \tau_p g$. From this a characteristic cluster length can also be estimated *a priori* as $\mathcal{L} = \tau_p^2 g$. To ensure the hydrodynamics are independent of the domain size, the simulation configurations are equal or larger than Case 4 reported in Capecelatro et al. [98].

To sample the parameter space typical of turbulent fluidized bed reactors [171], the mean particle-phase volume fraction is varied from $0.001 \leq \langle \varepsilon_p \rangle \leq 0.05$ and the Archimedes number is varied from $1.8 \leq \mathrm{Ar} \leq 18.0$ by adjusting gravity. Due to the large density ratios under consideration, the mean mass loading ranges from $\mathcal{O}(10)$–$\mathcal{O}(10^2)$, and consequently two-way coupling between the phases is expected to be important. Here, angled brackets denote both a spatial and a temporal average

(since the flow under consideration is triply periodic and statistically stationary in time). A list of relevant non-dimensional numbers and other important simulation parameters are summarized in table 4.1.

4.2.2 Eulerian–Lagrangian training data

The training data presented in this section were generated using the Eulerian–Lagrangian approach previously described in Sections2.2.3-2.2.5, with the heat and mass transfer terms treated as null. The simulations were initialized with a random distribution of particles and run for approximately $100\tau_p$ until the flow reached a statistically stationary state. At this point statistics are accumulated over $50\tau_p$. Instantaneous snapshots of the streamwise fluid velocity and particle position of each case at steady state are shown in figure 4.1. It can immediately be seen that clusters of particles are generated and entrain the fluid downward. As a consequence of the frame of reference under consideration, the fluid flows upward in regions void of particles. Clusters are seen to become more distinct with increasing $\langle\varepsilon_p\rangle$. The effect of Ar on the flow field is less noticeable. As shown in table 4.2, the standard deviation in volume fraction fluctuations $\langle\varepsilon_p'^2\rangle^{1/2}$ increases with increasing Ar, with $\varepsilon_p' = \varepsilon_p - \langle\varepsilon_p\rangle$, indicating enhanced clustering. Perhaps less obvious, the volume fraction fluctuations normalized by $\langle\varepsilon_p\rangle$ are maximum for the intermediate volume fraction case ($\langle\varepsilon_p\rangle = 0.025$).

Because turbulence in CIT is driven by two-way coupling with the particle phase, there exist few characteristic scales that can be calculated *a priori*. However, the Taylor Reynolds number $Re_\lambda = u_{rms}\lambda/\nu_f$ and the Stokes number $St_\eta = \tau_p/\tau_\eta$, both of which must be computed *a posteriori*, provide insight on the resulting turbulence. Here u_{rms} is the average root-mean-square velocity, $\lambda = \sqrt{15\nu_f/\varepsilon_f}u_{rms}$ is the Taylor micro-scale with ε_f the viscous dissipation rate and $\tau_\eta = \sqrt{\nu_f/\varepsilon_f}$ is the Kolmogorov time scale. The values of these quantities for all nine cases are reported in table 4.2.

	$\langle\varepsilon_p\rangle$	$\frac{\sqrt{\langle\varepsilon_p'^2\rangle}}{\langle\varepsilon_p\rangle}$	$\frac{\langle u_f'''^2\rangle_f}{k_f}$	$\frac{\langle v_f'''^2\rangle_f}{k_f}$	$\frac{\langle u_p''^2\rangle_p}{k_p}$	$\frac{\langle v_p''^2\rangle_p}{k_p}$	$\frac{\langle u_p\rangle_p}{\mathcal{V}_0}$	$\frac{\left(\langle u_f\rangle_p+\tau_p^\star g\right)}{\mathcal{V}_0}$	St_η	Re_λ
	0.001	0.63	1.49	0.25	1.48	0.26	1.87	1.53	0.63	6.90
Ar= 1.8	0.0255	0.76	1.58	0.21	1.61	0.20	2.56	2.39	2.57	4.10
	0.05	0.74	1.51	0.24	1.52	0.24	2.46	2.34	2.97	3.29
	0.001	0.71	1.68	0.16	1.67	0.16	1.63	1.36	1.55	12.54
Ar= 5.4	0.0225	0.87	1.61	0.19	1.63	0.19	2.28	2.21	5.18	10.23
	0.05	0.84	1.56	0.22	1.57	0.21	2.12	2.07	6.29	8.76
	0.001	0.72	1.76	0.12	1.76	0.12	1.26	1.15	3.43	24.14
Ar= 18.0	0.0255	1.00	1.70	0.15	1.73	0.14	1.86	1.91	9.92	32.54
	0.05	0.98	1.64	0.18	1.68	0.16	1.81	1.83	12.97	35.52

Table 4.2: Statistically stationary EL quantities for all nine training cases.

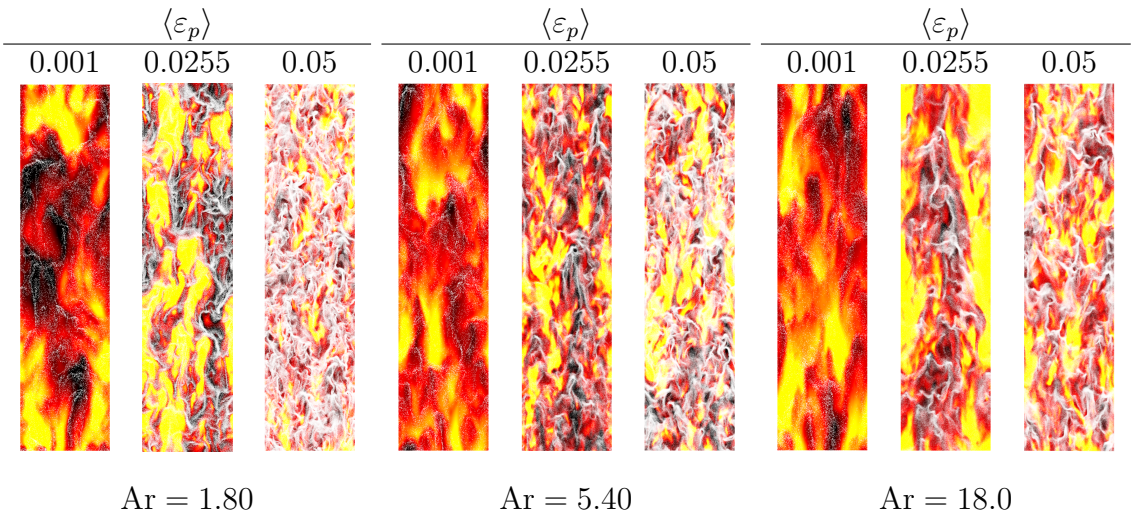


Figure 4.1: Instantaneous snapshots of fully developed CIT at statistical steady state. A slice at the centreline in the x–y plane is shown, with particle position (white) and normalized vertical fluid velocity $u_f/\mathcal{V}_0$ (colour).

It can be seen that high volume fractions correspond to larger Stokes numbers and lower values of Re_λ. Both St_η and Re_λ tend to increase when larger body forces are applied (i.e. larger Ar). Because fluid velocity fluctuations are generated by two-way coupling, $Re_\lambda > 0$ only when $\langle\varepsilon_p\rangle > 0$. That said, Re_λ is seen to decrease with increasing $\langle\varepsilon_p\rangle$ for $Ar = 1.8$ and 5.4. This is likely due to increased dissipation through drag exchange (see table 4.3). However, this reduction is less dramatic as Ar increases, and Re_λ is seen to be approximately constant at $Ar = 18$.

4.3 Phase-averaged equations

In this section, we present the phase-averaged flow equations in which we seek to model the unclosed terms that arise. This system of equations have been previously derived [41] and is extended here to take into account nonlinear drag effects due to F in Eq. (2.23).

Phase averaging (PA) is analogous to Favre averaging of variable-density flows and is denoted by $\langle(\cdot)\rangle_p = \langle\varepsilon_p(\cdot)\rangle/\langle\varepsilon_p\rangle$. Fluctuations about the PA particle velocity are expressed as $\boldsymbol{u}_p'' = \boldsymbol{u}_p(\boldsymbol{x},t) - \langle\boldsymbol{u}_p\rangle_p$, with $\langle\boldsymbol{u}_p''\rangle_p = 0$. This gives rise to the

PA particle-phase turbulent kinetic energy (TKE), $k_p = \langle \boldsymbol{u}_p'' \cdot \boldsymbol{u}_p'' \rangle_p / 2$. Here, $\boldsymbol{u}_p$ is the particle-phase velocity in an Eulerian frame of reference. It should be noted that $\langle \boldsymbol{u}_p \rangle_p$ is equivalent to the average particle velocity $\langle \boldsymbol{v}_p \rangle$ (with angled brackets here used to represent a particle average). Thus, $\langle u_p \rangle_p$ will be used throughout to characterize the mean settling velocity of the particle phase. In a similar fashion, the PA operator in the fluid phase is defined as $\langle (\cdot) \rangle_f = \langle \varepsilon_f (\cdot) \rangle / \langle \varepsilon_f \rangle$. Fluctuations about the PA fluid velocity are given by $\boldsymbol{u}_f''' = \boldsymbol{u}_f(\boldsymbol{x}, t) - \langle \boldsymbol{u}_f \rangle_f$. With this, the fluid-phase TKE is given by $k_f = \langle \boldsymbol{u}_f''' \cdot \boldsymbol{u}_f''' \rangle_f / 2$.

For the statistically stationary and homogeneous flows considered herein, continuity implies $\langle \varepsilon_f \rangle$ is constant and the fluid-phase momentum equation reduces to $\langle \boldsymbol{u}_f \rangle_f = 0$. In the particle phase, the only non-zero component of the averaged momentum equation is in the gravity-aligned direction (the x-direction in this case):

$$\frac{\partial \langle u_p \rangle_p}{\partial t} = \frac{1}{\tau_p^\star} \left(\langle u_f \rangle_p - \langle u_p \rangle_p \right) + \frac{1}{\rho_p} \left(\left\langle \frac{\partial \sigma_{f,xi}}{\partial x_i} \right\rangle_p - \left\langle \frac{\partial p_f}{\partial x} \right\rangle_p \right) + g, \tag{4.3}$$

noting that for gas–solid flows, the terms involving $\sigma_{f,xi}$ and p_f are small enough to be neglected [41]. This implies that at steady state, $\langle u_p \rangle_p \approx \langle u_f \rangle_p + \tau_p^\star g$. Here, we incorporate the nonlinearities associated with drag in $\tau_p^\star = \tau_p / \langle F_d \rangle_p$, where $\langle F_d \rangle_p (\langle \varepsilon_f \rangle, \langle \mathrm{Re}_p \rangle)$ is the nonlinear drag correction [111] defined using averaged flow arguments. This definition does not include the dependencies on drag covariance terms (i.e. $\langle u_f''' F_d'' \rangle_p$ and $\langle u_p'' F_d''' \rangle_p$), however, as shown in table 4.2, these terms have negligible contributions when describing particle settling, $\langle u_p \rangle_p$, and are thus neglected.

The transport equations for the fluid-phase Reynolds stresses can be reduced to two unique, non-zero components. In the streamwise direction this equation is given

as

$$\frac{1}{2}\frac{\partial\langle u_f'''^2\rangle_f}{\partial t} = \underbrace{\frac{1}{\rho_f}\left\langle p_f\frac{\partial u_f'''}{\partial x}\right\rangle}_{\text{pressure strain (PS)}} - \underbrace{\frac{1}{\rho_f}\left\langle \sigma_{f,1i}\frac{\partial u_f'''}{\partial x_i}\right\rangle}_{\text{viscous dissipation (VD)}} + \underbrace{\frac{\varphi}{\tau_p^\star}\left(\langle u_f'''u_p''\rangle_p - \langle u_f'''^2\rangle_p\right)}_{\text{drag exchange (DE)}} + \underbrace{\frac{\varphi}{\tau_p^\star}\langle u_f'''\rangle_p\langle u_p\rangle_p}_{\text{drag production (DP)}} + \underbrace{\frac{\varphi}{\rho_p}\left\langle u_f'''\frac{\partial p_f'}{\partial x}\right\rangle_p}_{\text{pressure exchange (PE)}} - \underbrace{\frac{\varphi}{\rho_p}\left\langle u_f'''\frac{\partial \sigma_{f,1i}'}{\partial x_i}\right\rangle_p}_{\text{viscous exchange (VE)}}. \tag{4.4}$$

Similarly, both cross-stream equations are given as

$$\frac{1}{2}\frac{\partial\langle v_f'''^2\rangle_f}{\partial t} = \underbrace{\frac{1}{\rho_f}\left\langle p_f\frac{\partial v_f'''}{\partial y}\right\rangle}_{\text{pressure strain (PS)}} - \underbrace{\frac{1}{\rho_f}\left\langle \sigma_{f,2i}\frac{\partial v_f'''}{\partial x_i}\right\rangle}_{\text{viscous dissipation (VD)}} + \underbrace{\frac{\varphi}{\tau_p^\star}\left(\langle v_f'''v_p''\rangle_p - \langle v_f'''^2\rangle_p\right)}_{\text{drag exchange (DE)}} + \underbrace{\frac{\varphi}{\rho_p}\left\langle v_f'''\frac{\partial p_f'}{\partial y}\right\rangle_p}_{\text{pressure exchange (PE)}} - \underbrace{\frac{\varphi}{\rho_p}\left\langle v_f'''\frac{\partial \sigma_{f,2i}'}{\partial x_i}\right\rangle_p}_{\text{viscous exchange (VE)}}, \tag{4.5}$$

where the drag production term no longer appears, since it is a gravity-driven phenomenon.

Due to the homogeneity of the flow and symmetry in the directions perpendicular to gravity (y and z directions in this configuration), the unique, non-zero PA Reynolds-stress transport equations in the particle phase are given as

$$\frac{1}{2}\frac{\partial\langle u_p''^2\rangle_p}{\partial t} = \underbrace{\left\langle \Theta\frac{\partial u_p''}{\partial x}\right\rangle_p}_{\text{pressure strain}} - \underbrace{\left\langle \sigma_{p,1i}\frac{\partial u_p''}{\partial x_i}\right\rangle_p}_{\text{viscous dissipation}} + \underbrace{\frac{1}{\tau_p^\star}\left(\langle u_f'''u_p''\rangle_p - \langle u_p''^2\rangle_p\right)}_{\text{drag exchange}} + \underbrace{\frac{1}{\rho_p}\left\langle u_p''\frac{\partial \sigma_{f,1i}'}{\partial x_i}\right\rangle_p}_{\text{viscous exchange}} - \underbrace{\frac{1}{\rho_p}\left\langle u_p''\frac{\partial p_f'}{\partial x}\right\rangle_p}_{\text{pressure exchange}}. \tag{4.6}$$

Similarly, the cross-gravity equations are both (due to symmetry and homogeneity)

given as

$$\frac{1}{2}\frac{\partial\langle v_p''^2\rangle_p}{\partial t} = \underbrace{\left\langle \Theta\frac{\partial v_p''}{\partial y}\right\rangle_p}_{\text{pressure strain}} - \underbrace{\left\langle \sigma_{p,2i}\frac{\partial v_p''}{\partial x_i}\right\rangle_p}_{\text{viscous dissipation}} + \underbrace{\frac{1}{\tau_p^\star}\left(\langle v_f''' v_p''\rangle_p - \langle v_p''^2\rangle_p\right)}_{\text{drag exchange}}$$
$$\underbrace{\frac{1}{\rho_p}\left\langle v_p''\frac{\partial \sigma_{f,2i}'}{\partial y_i}\right\rangle_p}_{\text{viscous exchange}} - \underbrace{\frac{1}{\rho_p}\left\langle v_p''\frac{\partial p_f'}{\partial y}\right\rangle_p}_{\text{pressure exchange}}, \tag{4.7}$$

where Θ and σ_p are the granular temperature and the particle-phase viscous stress tensor, respectively [41].

Due to the high density ratio in gas-solid flows, pressure exchange (PE) and viscous exchange (VE) are often negligible [41]. Therefore, the overall kinetic energy balance for CIT includes production due to drag (DP), pressure strain (PS), viscous dissipation (VD) and drag exchange (DE). As in single-phase turbulence, VD results from the resolved small-scale velocity fluctuations in the fluid phase. In contrast, the drag exchange terms involve (i) DE_2: viscous dissipation of unresolved fluid velocity fluctuations (in the viscous boundary layers around individual particles) and (ii) DE_1: energy transferred to the particles at the fluid-particle interface. Sundaram and Collins [172] showed that the unresolved dissipation arises in the point-particle model due to the difference between the fluid velocity at the particle location and the particle velocity. Our previous work [41] showed that the relative contribution of the interphase energy exchange ($\mathrm{De}_1/(\mathrm{DE}_1+\mathrm{DE}_2)$) is approximately 22%. Further, the ratio of resolved to unresolved viscous dissipation was found to be less than 6% and decreases with increasing Ar. Therefore we do not expect the unresolved viscous dissipation not captured in the Eulerian–Lagrangian model to impact the overall balance.

4.4 Closure modelling

4.4.1 Sparse regression with embedded invariance

The focus of this section is modelling the unclosed terms that appear in the fluid-phase Reynolds-stress equations (6.5) and (4.5). The data used to inform these closures, as discussed in Section 4.2.2, is averaged after the flow has become statistically stationary in time. These values are summarized in the table 4.3. In the streamwise direction, drag production (DP) is mostly balanced by drag exchange (DE). Pressure strain (PS) and viscous dissipation (VD) contain fluid-phase residual contributions, while pressure exchange (PE) and viscous exchange (VE) contain contributions from both phases. These terms are small compared to DP and DE, but are not negligible in general. In the cross-stream direction, DE is mostly balanced by PS.

Each unclosed term is considered individually and models are learned using the sparse regression methodology described in Beetham and Capecelatro [58] and summarized here. In this method, it is postulated that any tensor quantity $\mathbb{D}$ can be modelled using an invariant tensor basis, $\mathbb{T}$, and a set of ideal, sparse coefficients, $\hat{\beta}$,

$$\mathbb{D} = \mathbb{T}\hat{\beta}. \tag{4.8}$$

The ideal coefficients are determined by solving the optimization problem

$$\hat{\beta} = \min_{\beta} ||\mathbb{D} - \mathbb{T}\beta||_2^2 + \lambda||\beta||_1, \tag{4.9}$$

where β is a vector of coefficients that varies depending upon the choice of a user-specified sparsity parameter, λ and $||\cdot||_2^2$ and $||\cdot||_1$ represent the L-2 and L-1 norms, respectively. In the case of single-phase turbulence, this methodology can be used readily with previously derived minimally invariant basis sets [173]. It is helpful to note that sparse regression is openly available in several software packages, including

streamwise direction							
	$\langle\varepsilon_p\rangle$	PS	VD	DP	DE	VE	PE
	0.0100	-0.004	-0.007	0.03	-0.02	0	0
Ar = 1.8	0.0255	-0.189	-0.122	2.49	-2.75	-0.01	0
	0.5000	-0.406	-0.162	6.04	-6.87	-0.04	0
	0.0100	-0.021	-0.049	0.15	-0.138	0	0
Ar = 5.4	0.0255	-1.201	-0.482	15.18	-15.71	-0.06	0
	0.5000	-2.680	-0.709	39.67	-43.00	-0.29	0
	0.0100	-0.106	-0.264	0.81	-0.732	0	0
Ar = 18.0	0.0255	-9.988	-2.097	129.99	-131.1	-0.83	0
	0.5000	-22.056	-3.455	317.84	-329.3	-3.32	0

cross-stream directions						
	$\langle\varepsilon_p\rangle$	PS	VD	DE	VE	PE
	0.0100	0.002	-0.0004	-0.002	0	0
Ar = 1.8	0.0255	0.096	-0.0079	-0.171	0.002	0
	0.5000	0.219	-0.0140	-0.455	0.012	0.002
	0.0100	0.011	-0.0016	-0.01	0	0
Ar = 5.4	0.0225	0.616	-0.0293	-0.83	0.017	0.002
	0.5000	1.425	-0.0537	-2.323	0.083	0.007
	0.0100	0.053	-0.0052	-0.047	0	0
Ar = 18.0	0.0255	5.144	-0.1102	-5.686	0.183	0.009
	0.5000	12.020	-0.2305	-17.637	0.672	0.051

Table 4.3: Averaged terms for each contribution in the fluid-phase Reynolds-stress transport equations (6.5) and (4.5).

(1) Particle-phase anisotropic stress tensor	$\hat{\mathbb{R}}_p = \frac{\langle \boldsymbol{u}_p''' \boldsymbol{u}_p''' \rangle}{2k_p} - \frac{1}{3}\mathbb{I}$
(2) Fluid-phase anisotropic stress tensor	$\hat{\mathbb{R}}_f = \frac{\langle \boldsymbol{u}_f''' \boldsymbol{u}_f''' \rangle_f}{2k_f} - \frac{1}{3}\mathbb{I}$
(3) Slip velocity tensor	$\hat{\mathbb{U}}_r = \frac{\boldsymbol{U}_r}{\mathrm{tr}(\boldsymbol{U}_r)} - \frac{1}{3}\mathbb{I},$

Table 4.4: Second-order, symmetric, deviatoric tensors available to the multiphase RANS equations for modelling.

PySINDy [174]. However, to date an analogous basis has not yet been identified for multiphase flows. Due to the relative simplicity of the system under study (i.e. symmetry, homogeneity and stationarity), the parameters that may contribute to such a basis are limited to three tensors: the fluid-phase Reynolds stress anisotropy tensor, $\hat{\mathbb{R}}_f$, the particle-phase Reynolds stress anisotropy tensor $\hat{\mathbb{R}}_p$, and a the mean slip tensor, $\hat{\mathbb{U}}_r$ (see table 4.4). The mean slip tensor is defined as $\boldsymbol{U}_r = \boldsymbol{u}_r \otimes \boldsymbol{u}_r$, where $\boldsymbol{u}_r = \langle \boldsymbol{u}_p \rangle_p - \langle \boldsymbol{u}_f \rangle_f$ is the slip velocity vector. An important property of this vector is that in fully developed CIT it is always aligned with the direction of the body forcing (in this case gravity).

Because the sparse regression methodology postulates the model to be a linear combination of the basis tensors, this implies that the basis tensors must take on the same properties as the quantity to be modelled. The four terms under consideration here are all symmetric and thus the basis tensors must also be symmetric. The three tensor quantities shown in table 4.4 are used in order to formulate a minimally invariant basis by following the procedure described in Spencer and Rivlin [16]. This set of tensors, along with six scalar invariants, denoted $\mathcal{S}^{(i)}$, by definition can exactly describe the Eulerian–Lagrangian data. In the context of the sparse regression methodology, the ideal coefficients $\hat{\beta}$ may be constants or nonlinear functions of the scalar invariants, $\mathcal{S}^{(i)}$.

SectionubsectionResults and discussion Using the set of basis tensors defined in Section 4.4.1, the sparse regression methodology is employed to identify closures for the terms appearing in the fluid-phase Reynolds-stress equations (6.5) and (4.5),

$\mathcal{T}^{(1)} = \mathbb{I}$	$\mathcal{T}^{(2)} = \hat{\mathbb{U}}_r$	
$\mathcal{T}^{(3)} = \hat{\mathbb{U}}_r^2$	$\mathcal{T}^{(4)} = \left(\hat{\mathbb{U}}_r\hat{\mathbb{R}}_f\right)^\dagger$	
$\mathcal{T}^{(5)} = \left(\hat{\mathbb{U}}_r^2\hat{\mathbb{R}}_f\right)^\dagger$	$\mathcal{T}^{(6)} = \left(\hat{\mathbb{U}}_r^2\hat{\mathbb{R}}_f^2\right)^\dagger$	
$\mathcal{T}^{(7)} = \left(\hat{\mathbb{U}}_r\hat{\mathbb{R}}_f\hat{\mathbb{R}}_p\right)^\dagger$	$\mathcal{T}^{(8)} = \left(\hat{\mathbb{U}}_r^2\hat{\mathbb{R}}_f\hat{\mathbb{R}}_p\right)^\dagger$	
$\mathcal{T}^{(9)} = \left(\hat{\mathbb{R}}_f\hat{\mathbb{U}}_r^2\hat{\mathbb{R}}_p\right)^\dagger$	$\mathcal{T}^{(10)} = \left(\hat{\mathbb{U}}_r^2\hat{\mathbb{R}}_f^2\hat{\mathbb{R}}_p\right)^\dagger$	
$\mathcal{T}^{(11)} = \left(\hat{\mathbb{U}}_r\hat{\mathbb{R}}_f\hat{\mathbb{U}}_r^2\hat{\mathbb{R}}_p\right)^\dagger$	$\mathcal{T}^{(12)} = \left(\hat{\mathbb{U}}_r\hat{\mathbb{R}}_f\hat{\mathbb{R}}_p\hat{\mathbb{U}}_r^2\right)^\dagger$	
$\mathcal{T}^{(13)} = \hat{\mathbb{R}}_f$	$\mathcal{T}^{(14)} = \hat{\mathbb{R}}_f^2$	
$\mathcal{T}^{(15)} = \hat{\mathbb{R}}_p$	$\mathcal{T}^{(16)} = \hat{\mathbb{R}}_p^2$	
$\mathcal{T}^{(17)} = \left(\hat{\mathbb{U}}_r\hat{\mathbb{R}}_p\right)^\dagger$	$\mathcal{T}^{(18)} = \left(\hat{\mathbb{U}}_r^2\hat{\mathbb{R}}_p\right)^\dagger$	
$\mathcal{T}^{(19)} = \left(\hat{\mathbb{U}}_r\hat{\mathbb{R}}_p^2\right)^\dagger$	$\mathcal{T}^{(20)} = \left(\hat{\mathbb{U}}_r^2\hat{\mathbb{R}}_p^2\right)^\dagger$	
$\mathcal{T}^{(21)} = \left(\hat{\mathbb{R}}_f\hat{\mathbb{R}}_p\right)^\dagger$	$\mathcal{T}^{(22)} = \left(\hat{\mathbb{R}}_f^2\hat{\mathbb{R}}_p\right)^\dagger$	
$\mathcal{T}^{(23)} = \left(\hat{\mathbb{R}}_f\hat{\mathbb{R}}_p^2\right)^\dagger$	$\mathcal{T}^{(24)} = \left(\hat{\mathbb{R}}_f^2\hat{\mathbb{R}}_p^2\right)^\dagger$	
$\mathcal{S}^{(1)} = \mathrm{tr}\left(\hat{\mathbb{U}}_r\hat{\mathbb{R}}_f^2\hat{\mathbb{R}}_p^2\right)$	$\mathcal{S}^{(2)} = \mathrm{tr}\left(\hat{\mathbb{U}}_r\hat{\mathbb{R}}_f\hat{\mathbb{R}}_p^2\right)$	$\mathcal{S}^{(3)} = \mathrm{tr}\left(\hat{\mathbb{U}}_r\hat{\mathbb{R}}_f\hat{\mathbb{R}}_p\right)$
$\mathcal{S}^{(4)} = \mathrm{Ar}$	$\mathcal{S}^{(5)} = \varphi$	$\mathcal{S}^{(6)} = \langle\varepsilon_p\rangle$

Table 4.5: Minimally invariant set of basis tensors and associated scalar invariants. Here, $(\cdot)^\dagger = (\cdot) + (\cdot)^T$ denotes the tensor quantity added with its transpose.

based upon the Eulerian–Lagrangian data described in Section 4.2.2. Since flow data is homogeneous in all three spatial directions and we consider time-averaged data, each case is zero-dimensional (i.e. a single value).

As seen in table 4.3, the contributions from viscous and pressure exchange are either null, or relatively small even as mass loading is increased. For this reason, modelling efforts are directed toward the four remaining terms: drag production, pressure strain, viscous dissipation and drag exchange. Each is modelled separately, beginning with drag production as it is the sole source of fluid-phase turbulent kinetic energy in the absence of mean shear. As seen in Eq. (6.5), it is proportional to $\langle u_f''' \rangle_p$, which is zero in the absence of particle clusters.

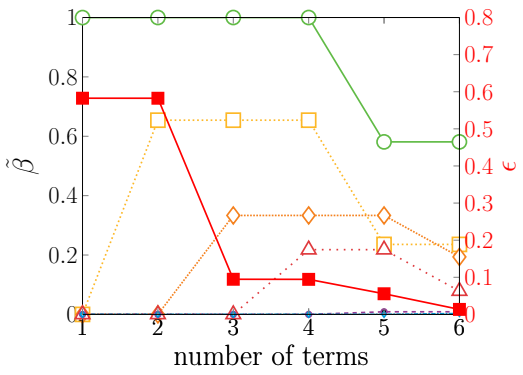


Figure 4.2: Normalized coefficients, $\tilde{\beta}$ (left axis) and associated model error, ϵ, (—■—, right axis) for drag production. The three-term and six-term models are described in equations 4.11 and 4.12, respectively. Terms 1–6, as denoted in equation 4.12, are represented as —●—, ···■···, ···◆···, ···▲···, - -●- -, -·-◆-·-, respectively. These colors also correspond with figure 4.4.

4.4.2 Drag production

As input to the sparse regression algorithm, drag production is non-dimensionalized using the square of the PA particle velocity, $\langle u_p \rangle_p^2$ and the drag time, τ_p. Because drag production is symmetric and also contains zero off-diagonal components, the basis set was restricted to only include terms that are functions of $\hat{\mathbb{U}}_r$ and $\mathbb{I}$, which also exhibit this property. While the Reynolds stresses have null off-diagonal components for this particular configuration, this does not hold in a general sense.

During optimization, as λ is decreased, additional terms are added to the learned model and model error decreases (see figure 4.2), where model error is defined as

$$\epsilon = \frac{||\mathbb{D} - \mathbb{T}\hat{\beta}||_2^2}{||\mathbb{D}||_2^2}. \tag{4.10}$$

In examining the relationship between model error and model complexity, we observe that a significant reduction in error is achieved with three model terms and error is drastically reduced when considering a six-term model. It is also notable that the most important terms to overall model performance appear in the models with lesser complexity and remain dominant as subsequent terms are added. This is indicated

by the behaviour of the normalized coefficients $\tilde{\beta}$, given as $\hat{\beta}^{(p)}/\max\hat{\beta}^{(1)}$, where p denotes the number of terms in the model.

The resultant learned models with three terms and six terms are given, respectively, as

$$\mathcal{R}^{\mathrm{DP}} = \frac{\langle u_p \rangle_p^2}{\tau_p}\left[\underbrace{1.11\varphi\hat{\mathbb{U}}_r}_{\text{term 1}} - \underbrace{0.73\varphi^{-2}\hat{\mathbb{U}}_r}_{\text{term 2}} + \underbrace{0.37\varphi\mathbb{I}}_{\text{term 3}}\right] \tag{4.11}$$

and

$$\mathcal{R}^{\mathrm{DP}} = \frac{\langle u_p \rangle_p^2}{\tau_p}\left[\underbrace{0.65\varphi\hat{\mathbb{U}}_r}_{\text{term 1}} - \underbrace{0.26\varphi^{-2}\hat{\mathbb{U}}_r}_{\text{term 2}} + \underbrace{0.22\varphi\mathbb{I}}_{\text{term 3}} - \underbrace{0.09\varphi^{-2}\mathbb{I}}_{\text{term 4}} + \underbrace{0.01\varphi^{2}\hat{\mathbb{U}}_r}_{\text{term 5}} + \underbrace{0.003\varphi^{2}\mathbb{I}}_{\text{term 6}}\right]. \tag{4.12}$$

To illustrate the interplay between model complexity and interaction, we consider the highly accurate, six-term model (see figures 4.3 (d)–4.3 (f) and Eq. 4.12) and the simpler, three-term model (see figures 4.3(a)–4.3(c) and Eq. 4.11). In comparing the performance of these models, we observe that the general scaling and spread of the data is captured reasonably well with the three-term model, but that the complexity added in the six-term model makes smaller adjustments that drive down model error. As shown in figure 4.4, the accuracy of the three-term model is primarily centered on the streamwise component of drag production (see figure 4.4 (a)); however, it over predicts the cross-stream components (see figure 4.4 (c)). The six-term model, in turn, reduces overall model error by more accurately describing both components; however, this is most pronounced in the cross-stream direction (see figures 4.4 (b) and 4.4 (d)).

In addition to discovering compact, algebraic models, sparse regression is also robust to sparse training data [58]. To illustrate this, a model was discovered using a sparse training dataset corresponding to $(\mathrm{Ar}, \langle \varepsilon_p \rangle) = [(1.8, 0.05), (5.4, 0.001), (18.0, 2.55)]$

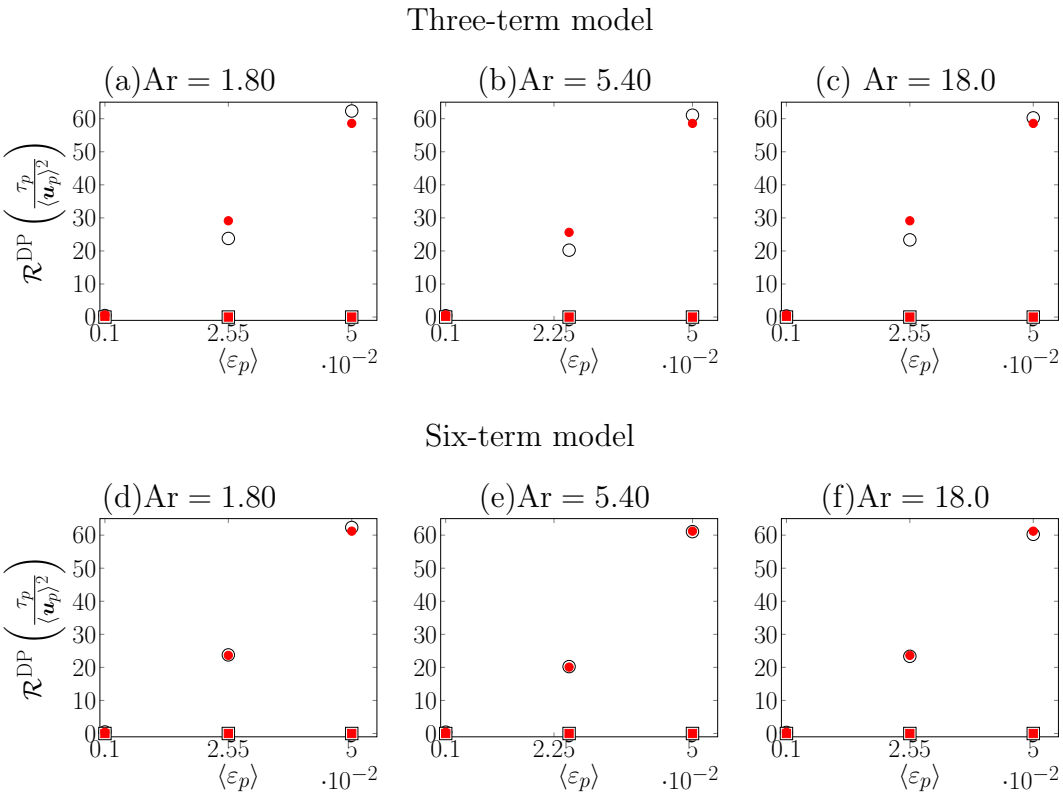


Figure 4.3: Drag production obtained from Eulerian–Lagrangian results (□, cross-stream component and ∘, streamwise components) and model prediction (■, cross-stream component and •, streamwise components). The model corresponds to Eq. 4.12 with $\lambda = 0.01$. The associated model error is $\epsilon = 0.01$.

and then tested using the remaining six cases. The resultant model is given as

$$\mathcal{R}^{\mathrm{DP}} = \frac{\langle \boldsymbol{u}_p \rangle^2}{\tau_p} \left[\left(0.36\varphi^{-1} + 0.05\varphi^2 - 4 \times 10^{-4}\varphi^3 \right) \hat{\mathbb{U}}_r + \left(0.01\varphi^2 + 0.21\varphi^{-1} \right) \mathbb{I} \right] \quad (4.13)$$

and shown compared with the trusted Eulerian–Lagrangian data in figures 4.5 (a)–4.5(c). It is notable that the sparsely trained model achieves reasonable accuracy using a subset of the available training points. While additional more challenging tests of the model are required and reserved for future work, this preliminary result may suggest model robustness to variations in Ar and particle volume fraction.

The remaining terms, pressure strain, viscous dissipation and drag exchange, ex-

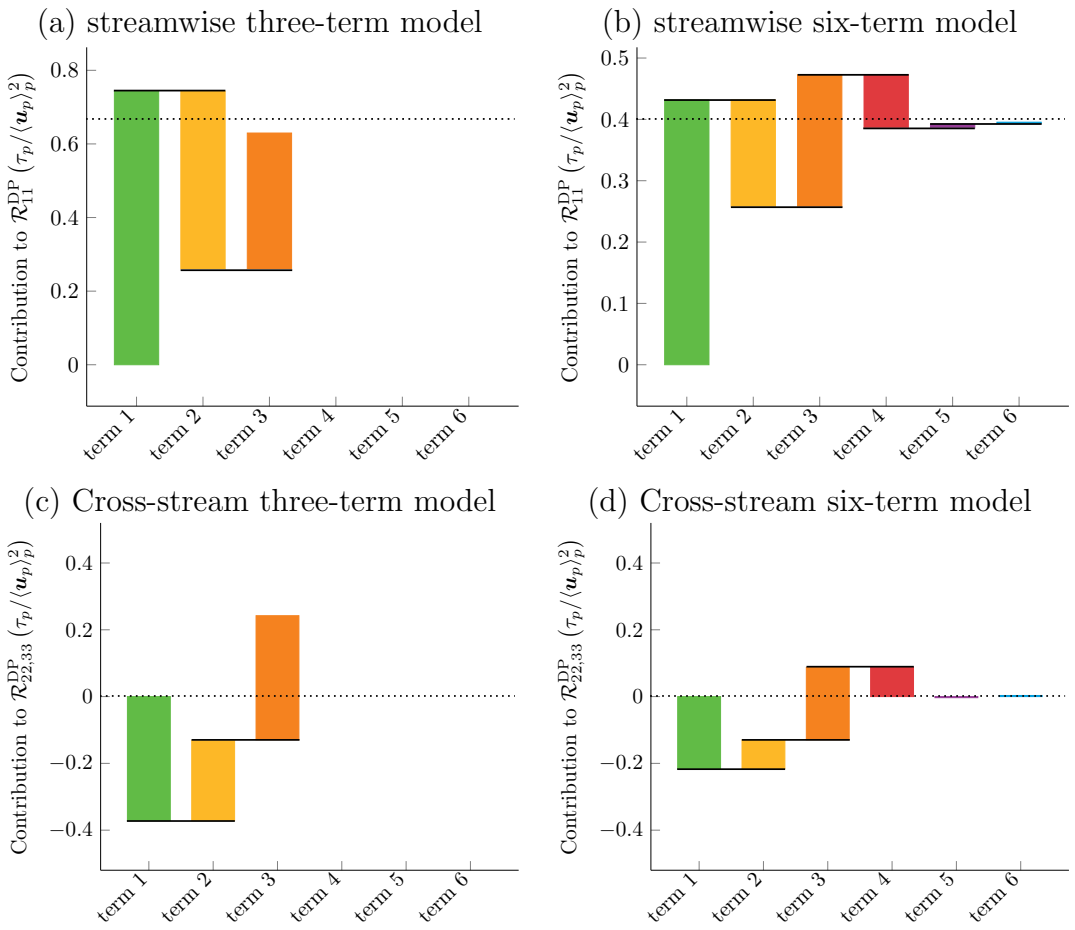


Figure 4.4: Term contributions for the streamwise component of drag production for the three-term (Eq. 4.11) and six-term (Eq. 4.12) models, shown for the case $Ar = 5.40$ and $\langle \varepsilon_p \rangle = 0.001$. Drag production obtained from the Eulerian–Lagrangian simulations is shown as the dotted line. Terms 1–6 are represented as ■, ■, ■, ■, ■ and ■, respectively.

hibit similar performance as drag production and are summarized here. All three terms are normalized by k_f/τ_p in order to ensure realizability in the Reynolds stresses. The sparse regression algorithm was given access to constant coefficients as well as coefficients dependent upon the invariants, $\mathcal{S}^{(i)}$, up to order three. It is notable that while complex nonlinear coefficients were accessible to the algorithm, they were ultimately not chosen.

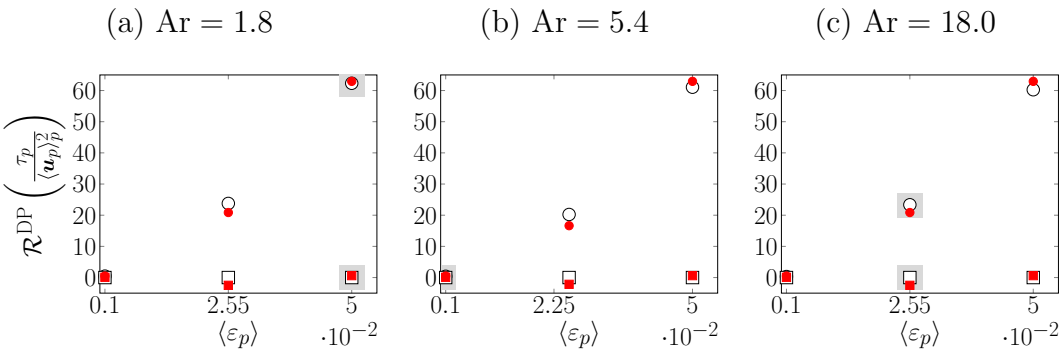


Figure 4.5: Model learned from sparse training data (denoted with grey shaded boxes). The training and testing error are 0.07 and 0.08, respectively. Using the convention from previous figures, Eulerian–Lagrangian results (○, streamwise component and □, cross-stream components) and model prediction (●, streamwise component and ■, cross-stream components). The sparsely trained model corresponds to Eq. 4.13.

4.4.2.1 Pressure strain and viscous diffusion

Pressure strain and viscous diffusion both redistribute turbulent kinetic energy throughout the fluid phase and are present in single-phase flows. The learned models are given as

$$\mathcal{R}^{\mathrm{PS}} = \varphi \frac{k_f}{\tau_p} \left[\langle \varepsilon_p \rangle \left(14.36 \hat{\mathbb{U}}_r - 22.65 \hat{\mathbb{R}}_p \right) + \langle \varepsilon_f \rangle \left(2.60 \hat{\mathbb{R}}_p - 2.72 \hat{\mathbb{R}}_f \right) \right] \tag{4.14}$$

and

$$\begin{aligned} \mathcal{R}^{\mathrm{VD}} = \frac{k_f}{\tau_p} \Big[& - 1.62 \hat{\mathbb{R}}_f \hat{\mathbb{R}}_p \hat{\mathbb{R}}_p + \varphi \langle \varepsilon_p \rangle \left(0.53 \mathbb{I} + 0.72 \hat{\mathbb{U}}_r - 3.14 \hat{\mathbb{R}}_f \hat{\mathbb{R}}_p \right) \\ & + \varphi \langle \varepsilon_f \rangle \left(0.74 \hat{\mathbb{R}}_f - 0.62 \hat{\mathbb{U}}_r \right) \hat{\mathbb{R}}_p \Big], \end{aligned} \tag{4.15}$$

respectively. The dominant terms important for capturing the behaviour of pressure strain across flow parameters are $\varphi \langle \varepsilon_p \rangle \hat{\mathbb{U}}_r$ and $\varphi \langle \varepsilon_p \rangle \hat{\mathbb{R}}_p$ and inclusion of only these two terms results in model error of $\epsilon = 0.15$.

In the case of viscous diffusion, a four-term model is learned in which the three terms that persist into the six-term model are $\varphi \langle \varepsilon_p \rangle \hat{\mathbb{U}}_r$, $\varphi \langle \varepsilon_p \rangle \hat{\mathbb{R}}_f \hat{\mathbb{R}}_p$ and $\hat{\mathbb{R}}_f \hat{\mathbb{R}}_p \hat{\mathbb{R}}_p$.

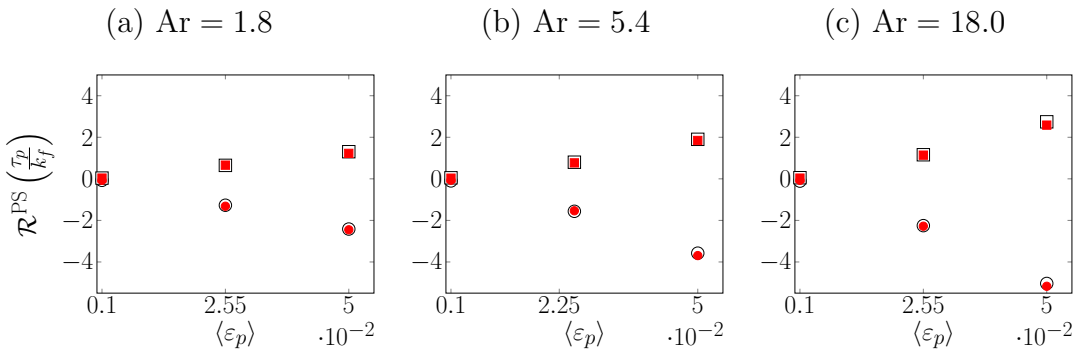


Figure 4.6: Pressure strain Eulerian–Lagrangian results (○, streamwise component and □, cross-stream components) and model prediction (●, streamwise component and ■, cross-stream components). Model corresponds to Eq. 4.14 and results from $\lambda = 0.3$. The associated model error is 0.04.

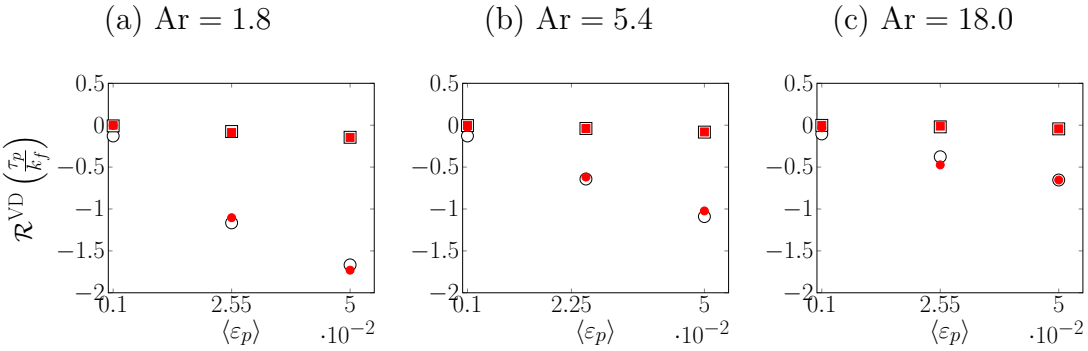


Figure 4.7: Viscous diffusion Eulerian–Lagrangian results (○, streamwise component and □, cross-stream components) and model prediction (●, streamwise component and ■, cross-stream components). Model corresponds to Eq. 4.15 and results from $\lambda = 0.2$. The associated model error is 0.07.

The fourth term, $\varphi\langle\varepsilon_p\rangle\hat{\mathbb{R}}_p$, is replaced by the three remaining terms that appear in Eq. 4.15. This reduces model error from 0.29 to 0.07, in a similar manner as described for drag production.

4.4.2.2 Drag exchange

Drag exchange describes the mechanism by which turbulent kinetic energy is partitioned between the phases. For this case, all terms in the model are of nearly equal importance. A four-term model is learned which excludes $\mathrm{Ar}\hat{\mathbb{U}}_r$ with an error of $\epsilon = 0.16$ as compared with the model error of $\epsilon = 0.15$ in the case of the five-term

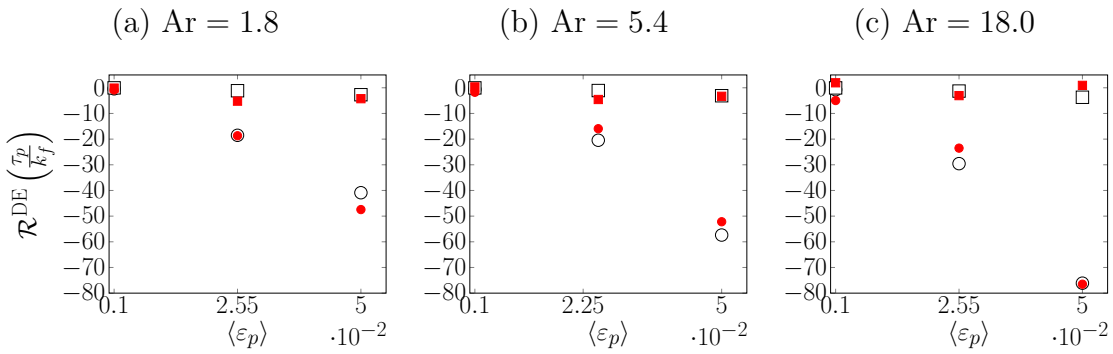


Figure 4.8: Drag exchange Eulerian–Lagrangian results (○, streamwise component and □, cross-stream components) and model prediction (●, streamwise component and ■, cross-stream components). Model corresponds to Eq. 4.16 and results from $\lambda = 0.006$. The associated model error is 0.15.

model. This is due to the minimal dependence of the data on Archimedes number.

$$\mathcal{R}^{\mathrm{DE}} = \varphi\frac{k_f}{\tau_p}\left[-0.36\varphi\mathbb{I} - \left(0.02\varphi^2\langle\varepsilon_f\rangle^2 + 0.38\mathrm{Ar}\right)\hat{\mathbb{U}}_r + 0.04\varphi^3\langle\varepsilon_f\rangle^3\hat{\mathbb{R}}_f\left(\hat{\mathbb{R}}_f - \hat{\mathbb{R}}_p\right)\hat{\mathbb{R}}_p\right] \tag{4.16}$$

For all of the terms considered, sparse regression is capable of uncovering models with model error to machine precision of zero (associated with $\lambda = 0$); however, these resultant models are substantially more complex and likely would not perform well outside the scope of training due to overfitting subtle nonlinearities. These models, for comparison, contain 18 terms for pressure strain, viscous diffusion and drag exchange, respectively, and 8 terms for drag production.

4.4.2.3 Particle-phase closures

The same procedure as described above was used to formulate closures for each of the terms appearing in the particle-phase Reynolds stress equations. The resultant closures are given as

$$\mathcal{R}^{\mathrm{PS}_p} = \varepsilon_p \left(1.89\hat{\mathbb{R}}_p + 2.12\hat{\mathbb{R}}_f\hat{\mathbb{R}}_p - 6.52\hat{\mathbb{R}}_f^2\hat{\mathbb{R}}_p \right) \tag{4.17}$$

$$\mathcal{R}^{\mathrm{VD}_p} = \varepsilon_p \left(-1.91\hat{\mathbb{R}}_p + 10.5\hat{\mathbb{U}}_r\hat{\mathbb{R}}_p - 11.78\hat{\mathbb{R}}_f\hat{\mathbb{R}}_p + 12.30\hat{\mathbb{R}}_f^2\hat{\mathbb{R}}_p \right) \tag{4.18}$$

$$\begin{aligned}\mathcal{R}^{\mathrm{DE}_p} = \varepsilon_f [&\left(-2.69 + 2.50\varphi\langle\varepsilon_f\rangle^3 - 1.85\varphi\langle\varepsilon_f\rangle\right)\mathbb{I} + \left(-4.79 + 4.08\varphi\langle\varepsilon_f\rangle^3 - 3.05\varphi\langle\varepsilon_f\rangle\right)\hat{\mathbb{U}}_r \\ &+ \left(9.10 - 5.56\varphi\langle\varepsilon_f\rangle\right)\hat{\mathbb{U}}_r\hat{\mathbb{R}}_p + \left(21.40\varphi\langle\varepsilon_f\rangle - 18.76\varphi\langle\varepsilon_f\rangle^3\right)\hat{\mathbb{R}}_f\hat{\mathbb{R}}_p]\end{aligned} \tag{4.19}$$

Here, the particle-phase terms are normalized by the dissipation of TKE in order to achieve appropriate scaling behavior with respect to time. The associated values of λ and model error for the pressure strain, viscous dissipation and drag exchange are given as $\lambda = (0.3, 0.3, 1)$ and $\epsilon = (0.08, 0.10, 0.03)$, respectively.

4.4.2.4 Summary of the multiphase RANS equations

In this section, we propose a complete set of transport equations for strongly-coupled gas-particle homogeneous flows, with a simpler set of equations for the transport of the total Reynolds Stresses. This allows for improved stability and robustness. The closures for individual balance terms are then useful to algebraically decompose the full Reynolds stresses into individual contributions. This system, along with the

particle momentum in equation 4.3, is given by

$$\frac{\partial \varepsilon_f}{\partial t} = \frac{1}{\tau_p^\star}\left[\frac{C_{\varepsilon_f}\mathcal{V}_0^2}{\tau_p^\star} - \varepsilon_f\right], \tag{4.20}$$

$$\frac{\partial \varepsilon_p}{\partial t} = \frac{1}{\tau_p^\star}\left[\frac{C_{\varepsilon_p}\mathcal{V}_0^2}{\tau_p^\star} - \varepsilon_p\right], \tag{4.21}$$

$$\frac{\partial \langle \boldsymbol{u}_f''' \boldsymbol{u}_f''' \rangle_f}{\partial t} = -\frac{1}{\tau_p^\star}\left(\langle \boldsymbol{u}_f''' \boldsymbol{u}_f''' \rangle_f + 0.5\langle \boldsymbol{u}_p'' \boldsymbol{u}_p'' \rangle_p\right) + \left(\frac{\mathcal{V}_0^2}{\tau_p^\star}\right)\left[(C_1 - C_2)\frac{\boldsymbol{U}_r}{\mathrm{tr}(\boldsymbol{U}_r)} + C_2\mathbb{I}\right], \tag{4.22}$$

$$\frac{\partial \langle \boldsymbol{u}_p'' \boldsymbol{u}_p'' \rangle_p}{\partial t} = -\frac{1}{\tau_p^\star}\left(\langle \boldsymbol{u}_p'' \boldsymbol{u}_p'' \rangle_p + 0.5\langle \boldsymbol{u}_f''' \boldsymbol{u}_f''' \rangle_f\right) + \left(\frac{\mathcal{V}_0^2}{\tau_p^\star}\right)\left[(C_3 - C_4)\frac{\boldsymbol{U}_r}{\mathrm{tr}(\boldsymbol{U}_r)} + C_4\mathbb{I}\right]. \tag{4.23}$$

Here, the coefficients, $C_{\varepsilon_f}, C_{\varepsilon_p}, C_1, C_2, C_3$ and C_4, depend upon flow parameters. These dependencies were learned using the sparse regression algorithm described previously and are all nonlinearly parameterized by the mass loading according to

$$C_{\varepsilon_f} = 0.44\langle\varphi\rangle - 0.05\langle\varphi\rangle^{1.5} - 0.21\langle\varphi\rangle^{-2}, \tag{4.24}$$

$$C_{\varepsilon_p} = 0.01\langle\varphi\rangle - 0.001\langle\varphi\rangle^{1.5} + 0.02\langle\varphi\rangle^{-2}, \tag{4.25}$$

$$C_1 = 1.40\langle\varphi\rangle - 0.18\langle\varphi\rangle^{1.5} + 2.17\langle\varphi\rangle^{-2}, \tag{4.26}$$

$$C_2 = 0.18\langle\varphi\rangle - 0.02\langle\varphi\rangle^{1.5} + 0.12\langle\varphi\rangle^{-2}, \tag{4.27}$$

$$C_3 = 1.20\langle\varphi\rangle - 0.16\langle\varphi\rangle^{1.5} + 2.03\langle\varphi\rangle^{-2}, \tag{4.28}$$

$$C_4 = 0.15\langle\varphi\rangle - 0.02\langle\varphi\rangle^{1.5} + 0.11\langle\varphi\rangle^{-2}. \tag{4.29}$$

The fluid drift velocity, defined as $\boldsymbol{u}_d = \langle \boldsymbol{u}_f \rangle_p - \langle \boldsymbol{u}_f \rangle_f$, requires modelling for the fluid velocity seen by the particles, $\langle \boldsymbol{u}_f \rangle_p$, as does the particle phase momentum

equation 4.3. The learned model for $\boldsymbol{u}_d$ is given as

$$\boldsymbol{u}_d = \frac{\sqrt{\langle \varepsilon_p'^2 \rangle}}{\langle \varepsilon_p \rangle} \left[-7.8\hat{\mathbb{R}}_f + 4.6\hat{\mathbb{R}}_p + 2.2 \right] \left(\langle \boldsymbol{u}_p \rangle_p - \langle \boldsymbol{u}_f \rangle_f \right) \tag{4.30}$$

using $\lambda = 1$, resulting in a model error of 0.06. Here, we impose that the model be scaled by the variance of the particle volume fraction, $\langle \varepsilon_p'^2 \rangle$, in order to ensure this quantity approaches zero in the dilute limit.

Finally, since the variance of particle volume fraction is not known *a priori*, an additional model was formulated for this quantity, using the same method described previously, given by

$$\frac{\sqrt{\langle \varepsilon_p'^2 \rangle}}{\langle \varepsilon_p \rangle} = 0.32\hat{\mathbb{R}}_p : \hat{\mathbb{R}}_p + \hat{\mathbb{U}}_r : \left(1.10\hat{\mathbb{R}}_f + 2.67\hat{\mathbb{R}}_p \langle \varepsilon_p \rangle^{1/2} \right). \tag{4.31}$$

4.4.3 Application to transient flow

To assess model realizability, it is imperative to evaluate the transient behavior that precedes the stationary state. To this end, we generate transient data by using the statistically stationary solution from the cases described in section 4.2 as an initial condition and instantaneously reverse the direction of gravity, i.e.

$$\boldsymbol{g} = \begin{cases} (-g, 0, 0), & \text{if } t < 0 \\ (g, 0, 0), & \text{if } t \geq 0 \end{cases}. \tag{4.32}$$

This strategy generates temporally evolving data and allows us to apply our models to a homogeneous flow, while relaxing the assumption of stationarity. Three cases corresponding to Ar, $\langle \varepsilon_p \rangle = [(1.8, 0.0255), (5.4, 0.05), (18, 0.001)]$ were simulated in order to probe the entire parameter space. The performance of the drag production model, in particular, is highlighted in figure 4.10 and compared against the transient

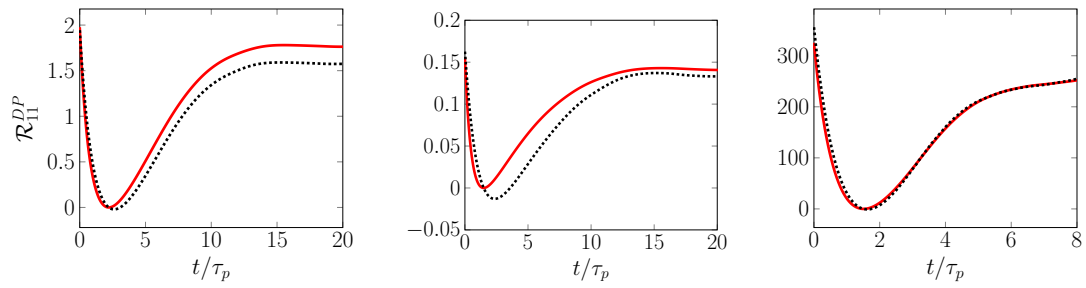


Figure 4.9: Temporal evolution of drag production obtained from the learned model 4.12 (——) and Euler–Lagrange data (.....) for CIT after gravity is reversed instantaneously.

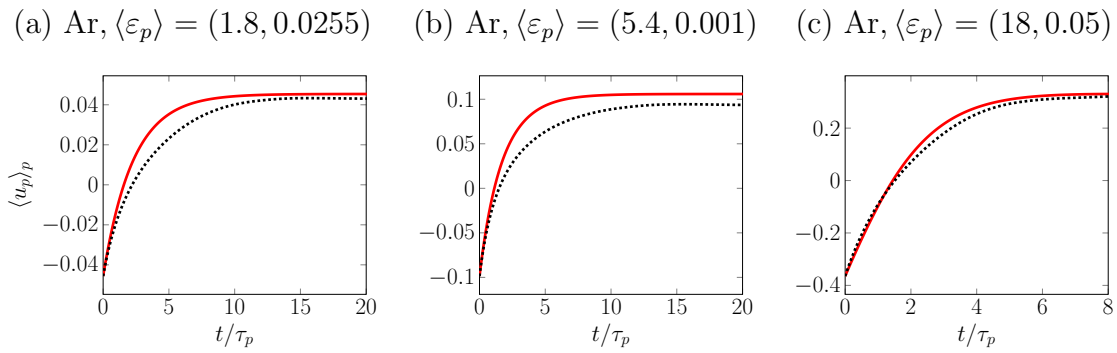


Figure 4.10: Temporal evolution of mean particle settling velocity obtained from the multiphase RANS equations (——) and Euler–Lagrange data (.....) for CIT after gravity is reversed instantaneously.

EL data for all three cases considered. Finally, the forward solution of the system of model equations presented is solved for these cases, and the predicted mean particle velocities are compared with the EL data.

The models described herein are successful for the parameter space studied and perform exceptionally well on transient data, despite being trained using stationary data. While care has been taken to ensure asymptotic agreement with the dilute limit, the robustness of the models applied to denser systems or non-homogeneous flows is left for future investigation.

4.5 Conclusions

In this Chapter, the multiphase RANS equations are presented for two-way coupled gas–solid flows. In this class of flows, the coupling between the phases spontaneously gives rise to coherent particle structures, which in turn generate and sustain turbulence in the carrier phase. This phenomenon has important engineering implications [4, 151, 56, 6] and makes the formulation of closure models that are predictive across scales and flow conditions challenging.

We apply a newly formulated modelling methodology, sparse regression with embedded form invariance [58], to highly-resolved Eulerian–Lagrangian data for fully-developed CIT. The benefits of this methodology as compared with Neural Networks, which have become increasingly popular, are (1) interpretability of the resultant closures, since they are in a closed algebraic formulation, (2) ease of dissemination to existing RANS solvers and (3) robustness to very sparse training sets. The dataset used for model development spans a range of flow parameters, specifically $\mathrm{Ar} = (1.8, 5.4, 18.0)$ and $\langle \varepsilon_p \rangle = (0.001, 0.0255, 0.05)$, in order to formulate models across a range of typical conditions.

Particular attention is paid to the closures for the four dominant unclosed terms that appear in the fluid-phase Reynolds stress equations – pressure strain, viscous diffusion, drag production and drag exchange. In applying the sparse regression method to each of these terms individually, we discover compact closures containing between four and six term that are accurate across the scope of training (model error ranges from 0.01 to 0.15). Because of the compact nature of the models developed and the nature of the sparse regression algorithm, we are able to assess the relative importance of each term and its role in reducing model error. Further, we demonstrate that even when training on a subset of the Eulerian–Lagrangian data, the methodology learns a model that remains accurate outside the scope of its training. Additionally, because of the compact, algebraic formulation of the method, resultant models are

accessible for interpretation and terms of greater physical significance are easily identified. These findings suggest that the sparse regression methodology holds promise for developing closures for more complicated multiphase flows, such as channel, duct or bubbly flows. Further, since nearly all flows of practical importance involve both walls and strong shear, future modelling work will focus on near-wall treatments and adaptation of the tensor basis to account for shear and other flow conditions which result in turbulence.

CHAPTER V

On the Thermal Entrance Length in Gas-Solid Flows

The dissipative nature of heat transfer relaxes thermal flows to an equilibrium state that is devoid of temperature gradients. The distance to reach an equilibrium temperature – the thermal entrance length – is a consequence of diffusion and mixing by convection. The presence of particles can modify the thermal entrance length due to interphase heat transfer and turbulence modulation by momentum coupling. In this work, Eulerian–Lagrangian simulations are utilized to probe the effect of solids heterogeneity (e.g., clustering) on the thermal entrance length. For the moderately dense systems considered here, clustering leads to a factor of 2–3 increase in the thermal entrance length, as compared to an uncorrelated (perfectly mixed) distribution of particles. The observed increase is found to be primarily due to the covariance between volume fraction and temperature fluctuations, referred to as the fluid drift temperature. Using scaling arguments and Gene Expression Programming, closure is obtained for this term in a one-dimensional averaged two-fluid equation and is shown to be accurate under a wide range of flow conditions.

5.1 Introduction

Internal flow exhibiting purely dissipative heat transfer exchanges heat with walls or its surroundings and its temperature profile relaxes to an equilibrium. The thermal entrance length, l_{th}, is defined as the length after which temperature gradients with respect to non-homogeneous directions vanish. Over the last several decades, the thermal entrance length has been studied extensively in the context of laminar and turbulent single-phase flows [see, e.g., 175, 176, 177, 178]. For laminar flow bounded by constant temperature walls, the entrance length may be estimated by [179]

$$l_{\text{th}}/D = 0.05 Re_D Pe, \tag{5.1}$$

where Re_D is the Reynolds number characterized by the vessel (pipe/duct) width or diameter, D, and Pr is the Prandtl number. When considering heat transfer in turbulent flows, the Nusselt number is used to assess when the flow has reached a fully-developed state. In particular, the entrance length is defined as the length after which the Nusselt number is within several percent (typically 1 to 5%) of the fully developed value and can be thought of as the thermal equivalent to a hydrodynamic boundary layer. While values across this range are used throughout the literature, Sparrow et al. [175] pointed out that 5% has more utility for comparison with experimental results, where achieving accuracy within 1 or 2% is challenging. Several models have been proposed for the Nusselt number in recent years, based upon the thermally evolving, turbulent pipe flow. Many draw upon the Dittus–Boelter correlation [180], given as $\text{Nu}= 0.023\text{Re}_D^{0.8}\text{Pr}^n$ where $n = 0.4$ for a heated fluid and 0.3 if it is cooled. Several other works also formulate dependencies upon the nondimensional length scale L/D [e.g., 178], where L is the pipe length.

While thermally evolving and wall-bounded, single-phase flows are of great importance (e.g., cooling systems for nuclear reactors, tube heat exchangers, etc.), many

applications of interest also contain a disperse phase that exchanges heat with the fluid. Of particular interest in this work are turbulent and thermally evolving *gas–solid* flows. This class of flows is pervasive in nature and industry, spanning applications from volcanic eruptions [181, 182, 183] to the storage of thermal energy [152, 184, 185, 186] and the upgrading of feedstock to usable fuels in circulating fluidized bed (CFB) reactors. In the case of CFB reactors, cool feedstock particles are fluidized with a hot gas, with the goal of mixing the phases in such a way that the hot, fluidizing gas exchanges heat with the particles, thereby initiating their devolatilization into usable fuels. In both experimental and computational studies, it has been observed that particles spontaneously organize into coherent structures (clusters), thereby reducing contact between the phases, impeding mixing and delaying heat transfer.

Early experimental work in the 1990s by Louge et al. [187], Ebert and Glicksman [188] showed that heat transfer between the particles in a fluidized bed and surrounding walls is increased by up to an order of magnitude when compared to single-phase turbulent flow. This increase in heat transfer was even more marked in denser regions of particles. In the context of a CFB reactor, [189] found that dilute suspensions of particles have the opposite effect and can impede heat transfer to the wall. In addition to these experimental works, several contemporary computational works have demonstrated that coherent structuring of particles may inhibit mixing between phases and detrimentally impact heat transfer [67, 151, 152, 56, 6]. This phenomenon has important implications for reactor design and efficiency, since reduced heat transfer performance impacts the thermochemical conversion rate. Despite the thermal entrance length's crucial role in properly sizing industrial unit operations, the effect of solids heterogeneity on this quantity remains largely unknown.

In the last decade, advancements in high performance computing has allowed for increased access to high-fidelity and large-scale computational studies of complex

multiphase flows. For example, Lei et al. [190] use progressive filtering of highly resolved simulations to formulate an improved model for the interphase heat transfer coefficient. Rauchenzauner and Schneiderbauer [191] derive a spatially averaged Euler–Euler model for heat transfer for wall-bounded, dense gas–solid flows by proposing a drift temperature that represents the fluid temperature fluctuations seen by the particles, the primary quantity of interest in the present work, and propose closures. Jofre et al. [192] studied heat transfer in irradiated turbulent dilute gas–solid flow using a two-step approach similar to the one undertaken in this work. They determined that the residence time and structure of particles play dominant roles in describing heat transfer. Another recent work [193] demonstrated that the pseudo-turbulent heat flux that arises in filtering the heat equation, is an important factor describing thermal properties. Yousefi et al. [194] employed particle-resolved direct numerical simulation to probe the heat transfer in particle-laden channels at moderate volume fractions, demonstrating that the turbulent heat flux dominates large scale thermal behavior.

While research on multiphase heat transfer is active and growing, the effect of particle heterogeneity on the thermal entrance length remains an open question. In this work, the thermal entrance length is examined via Eulerian–Lagrangian simulations by employing a two-step approach. First a moderately dense isothermal, gas-particle flow is simulated to generate realistic clustering. Next, the cold-flow simulations are fed into a statistically one-dimensional domain with a prescribed temperature difference between the phases at the inlet boundary. From these results, we quantify the effect of mean solids volume fraction, Péclet number and ratio of heat capacities on the thermal entrance length and propose scaling relations for both clustered and uniform gas–solid flows. We then derive the two-fluid heat equations, quantify which terms are responsible for deviations from an uncorrelated solids phase and propose a model for the dominant, unclosed term. This model is formulated using both scal-

ing arguments and Gene Expression Programming and is compatible with existing two-fluid theory.

5.2 System configuration

In this work, our goal is to examine the effect of realistic multiphase hydrodynamics on heat transfer and thermal entrance length. To do this, we use a two-step simulation setup representative of the fully-developed interior of a riser in a CFB reactor. A sketch of this configuration is outlined in Figure 5.1. Here, clustering behavior is established in an isothermal simulation, which then is fed into a thermally-evolving domain.

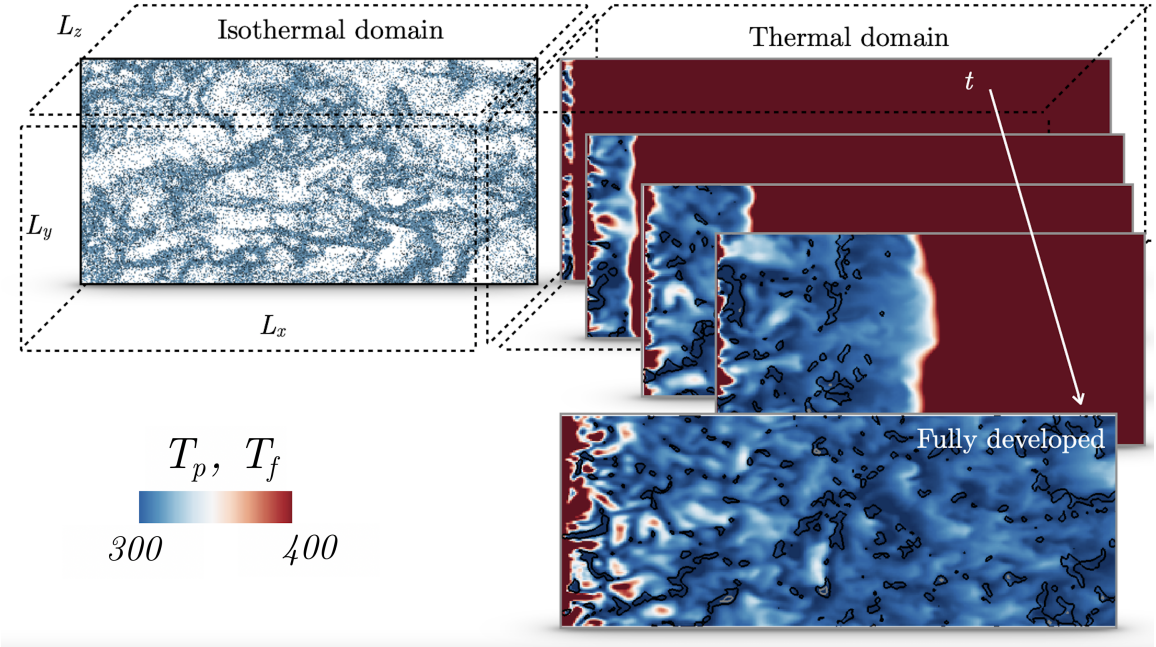


Figure 5.1: A fully developed configuration of particles (cold) and gas (hot) is injected into the thermal domain. Here, the initially cold particles are shown in the left pane. On the right, instantaneous snapshots of the fluid phase temperature is shown from an early time (top) to a fully developed period (bottom).

5.2.1 Isothermal simulations

Prior to simulating thermally-evolving two-phase flows, the hydrodynamics are established in a separate set of simulations. We consider three-dimensional, homogeneous, fully-developed gas–solid riser flow in the absence of heat transfer. In these simulations, N_p particles each with diameter d_p and density ρ_p are initially randomly distributed in a quiescent gas with density ρ_f and kinematic viscosity ν_f. A body force (gravity, g) drives the hydrodynamics and the mass flow rate is forced such that the mean fluid velocity is held at a fixed value, u_{bulk}, mimicking the flow conditions inside a riser. Here, u_{bulk} exceeds the anticipated particle settling velocity, and opposes the direction of gravity such that the particles are entrained in the fluid.

Due to strong coupling between the phases, the particles form dense clusters that generate correlation between the particle volume fraction, ε_p, and fluid velocity, u_f. When in a correlated (clustered) configuration, assemblies of particles experience enhanced settling, on the order of 2 to 3 times greater than the terminal velocity of an isolated particle, $\mathcal{V}_0$. This increased settling establishes a mean slip velocity between the phases that is not known *a priori* [see 40, 59, for more details].

In this configuration, relatively few non-dimensional groups arise. These include the Galileo number, $\text{Ga} = \sqrt{(\rho_p/\rho_f - 1)d_p^3 g}/\nu_f$; the bulk Reynolds number, $\text{Re}_{\text{bulk}} = u_{\text{bulk}} d_p/\nu_f$; and the mean mass loading $\varphi = \rho_p\langle\varepsilon_p\rangle/(\rho_f\langle\varepsilon_f\rangle)$. Here, $\varepsilon_f = 1 - \varepsilon_p$ is the fluid-phase volume fraction, and angled brackets denote an average in all three spatial directions and time. The parameters associated with the isothermal simulations are summarized in Table 5.1, where sets of values denote quantities that are varied in the simulations. Further details on the set up and flow physics of these simulations can be found in Beetham et al. [59].

	Dimensional quantities		
d_p	Particle diameter	90	[μm]
ρ_p	Particle density	1000	[kg/m^3]
ρ_f	Fluid density	1	[kg/m^3]
ν_f	Fluid viscosity	1.8×10^{-5}	[kg/m s]
u_{bulk}	Bulk fluid velocity	(0.42, 2.11, 2.95)	[m/s]
g	Gravity	0.8	[m/s^2]
τ_p	Stokes response time	0.025	[s]
	Non-dimensional Quantities		
N_p	Number of particles	(610,370, 15,564,442, 30,518,514)	
φ	Mass loading	(1.0, 26.2, 52.6)	
Ga	Galileo number	2.3	
$\langle\varepsilon_p\rangle$	Mean particle volume fraction	(0.001, 0.0255, 0.05)	
Re_{bulk}	Reynolds number	(2.1, 10.5, 14.7)	
	Computational quantities		
	Domain size ($W \times L \times L$)	$0.158 \times 0.038 \times 0.038$	[m]
	Grid size ($n_x \times n_y \times n_z$)	$512 \times 128 \times 128$	

Table 5.1: Summary of relevant parameters for the isothermal simulations.

5.2.2 Thermal simulations

Once a statistically stationary state is reached in time, a snapshot of the isothermal simulation is then fed into the thermal domain. This domain is initially comprised of fluid with heat capacity, $C_{p,f}$, and thermal diffusivity, κ_f. At the inlet, the fluid temperature is given a uniform value, $T_{f,0}$. In the spanwise directions, periodic boundary conditions are employed and the domain lengths match the isothermal simulation. In the streamwise direction, $y-z$ plane data is taken incrementally from the isothermal snapshot and fed in as an inlet condition at $x = 0$. Particles are assigned a uniform temperature, $T_{p,0} < T_{f,0}$, and heat capacity, $C_{p,p}$. After injection, two-way coupling drives the phases toward thermal equilibrium.

The thermal simulations introduce three additional relevant dimensionless groups: the Prandtl number, Pr= $C_{p,f}\nu_f\rho_f/\kappa_f$; the Péclet number, Pe= $d_p u_{\text{bulk}}\rho_f C_{p,f}/\kappa_f$; and the ratio of heat capacities, $\chi = C_{p,p}/C_{f,p}$. The parameters used in these simulations

are summarized in Table 5.2, where sets of values are provided for the quantities that are varied in the simulations. Using the riser of a CFB reactor as our motivation, we prescribe the inflow velocity, u_{bulk}, such that the resultant Péclet number corresponds to typical riser conditions [195]. Three values of χ are considered, corresponding to the heat capacities of sand [196], Zeolite 4A [197] (a catalyst used in the processing of conventional oil) and bagasse [96] (a woody pulp biproduct of the commercial processing of sugarcane commonly used in biomass pyrolysis).

	Particle-phase quantities		
$C_{p,p}$	Particle heat capacity	(840, 921 2300)	[J/kg K]
$T_{p,0}$	Initial particle temperature	300	[K]
	Fluid-phase Quantities		
$C_{p,f}$	Fluid heat capacity	1.013	[kJ/kg K]
$T_{f,0}$	Initial fluid temperature	400	[K]
κ_f	Fluid thermal conductivity	0.0334	[J/m s K]
	Non-dimensional quantities		
Pr	Prandtl number	0.7	
Pe	Péclet number	(1, 5, 7)	
χ	Ratio of heat capacities	(829, 909, 2270)	

Table 5.2: Summary of parameters for the thermally evolving simulations.

5.3 Computational methodology

The numerical simulations are solved in a volume-filtered Eulerian–Lagrangian framework for an incompressible viscous fluid with spherical, rigid particles undergoing heat exchange [81, 56]. The volume-filtered equations and details about the code used to solve them are described in Secs. 2.2.3-2.2.5. For the configurations under study here, the mass transfer and chemical kinetics are neglected for the thermal simulations and in the isothermal simulations, both heat and mass transfer are neglected.

5.4 Results

In this section, we summarize the results of the Euler–Lagrange simulations carried out using the setup and parameters discussed in Section 5.2 and the computational framework laid out in Section 5.3. We begin by reporting high level observations of both the cold-flow and thermal simulations, and then show profiles for the mean temperatures and quantify the thermal entrance length for each case. Finally, we propose scaling relations for the thermal entrance length corresponding to uncorrelated (uniform) particles and another that takes clustering into account.

5.4.1 Flow visualization

Three cold-flow (isothermal) simulations are performed using the parameters summarized in Table. 5.1. Instantaneous snapshots are shown in Figure 5.2. Beginning from an initially random distribution of particles, particles fall under gravity while the mean mass flow rate of the gas phase is held constant, allowing for a mean slip velocity between the phases to be established and a statistically stationary state to be reached after approximately $50\tau_p$. The degree of clustering is seen to vary as the volume fraction is increased. Dense suspensions of particles entrain the gas phase downward, resulting in so-called jet bypassing (high-speed upward flow in regions devoid of particles). At this point, the fluid-phase turbulent kinetic energy is produced by wakes behind clusters and shear layers at the edge of clusters, referred to as fully-developed cluster-induced turbulence (CIT) [40, 41].

In the thermal simulations, cool particles and hot gas from fully-developed CIT are injected into the thermal domain and heat transfer is enabled between the phases. As shown in Figure 5.3, the cool centers of dense clusters persist far into the domain and cool the surrounding fluid, while regions of dilute particles are heated more rapidly and have a minimal effect in cooling the fluid. This behavior is observed to be more dramatic for lower volume fraction and low Péclet number. Not surprisingly, as the

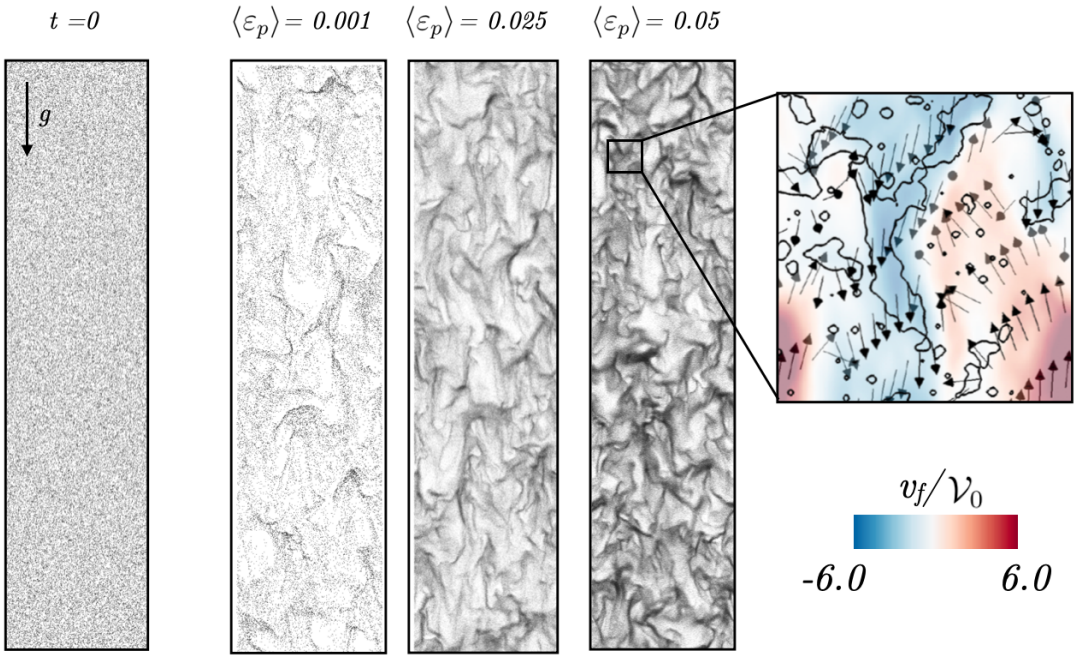


Figure 5.2: The isothermal simulations begin with an initially random distribution of particles (left) and evolve into a statistically stationary state characterized by clustering (middle 3 panels). These clusters generate and sustain turbulence in the fluid phase. Clusters entrain the fluid as they fall resulting in upflow in regions void of particles (right panel).

volume fraction is increased, the increase in mass loading of cold particles can more rapidly cool the surrounding gas, though hot spots still appear in regions devoid of particles. This behavior is shown in Figure 5.4.

5.4.2 The thermal entrance length

To quantify the thermal entrance length, we extend the definition from single-phase pipe flow to the configuration under consideration. Since both phases relax to an equilibrium temperature, a nondimensional thermal entrance length, l_{th}, can be defined as the location after which the difference between the mean temperatures is within 5% of the inlet temperature difference, or

$$l_{th} := \min\left(\hat{x} \in |\langle\theta_f\rangle - \langle\theta_p\rangle| \leq 0.05\right), \tag{5.2}$$

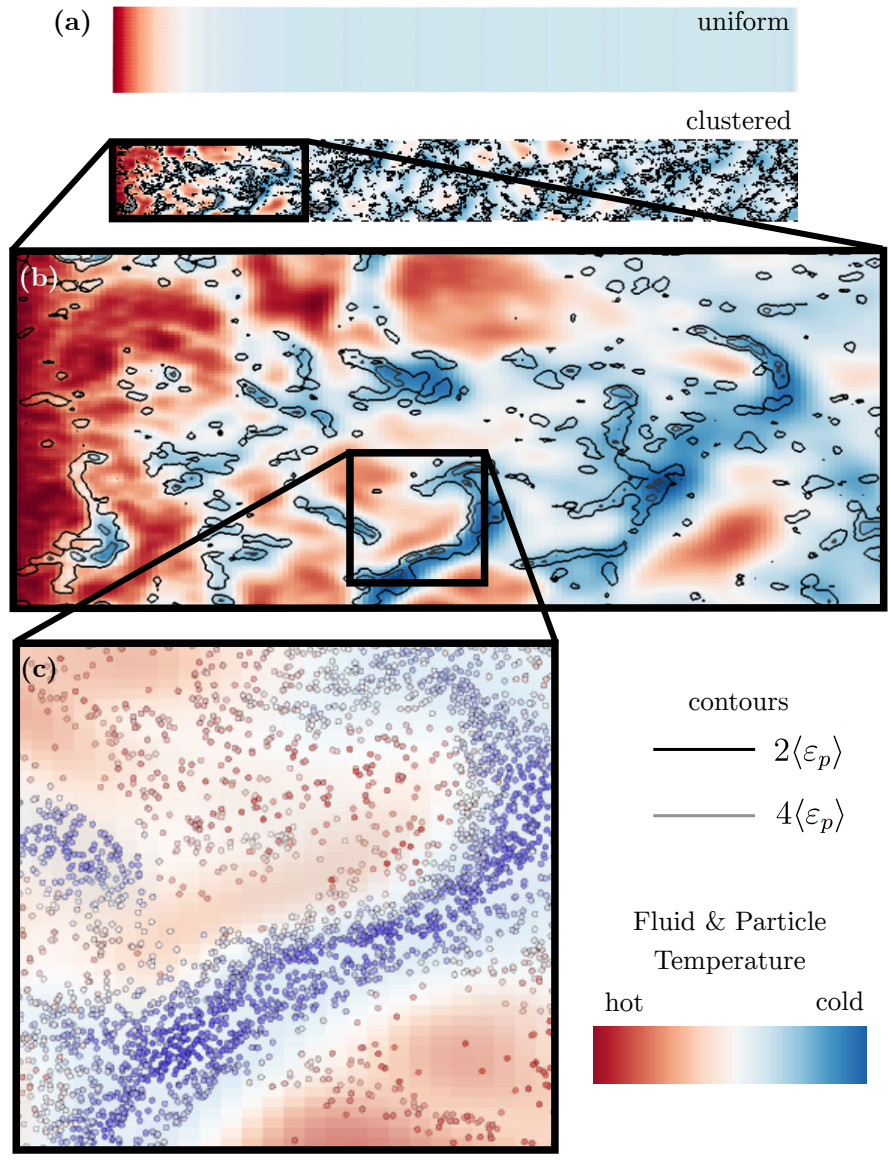


Figure 5.3: Hot (red) gas and cold (blue) particles are fed into a hot, quiescent thermal domain. From top to bottom: (a) When the particle phase is uncorrelated (uniformly distributed), the thermal entrance length is shorter as compared with a correlated (clustered) configuration of equal solid-phase volume fraction. (b) Clusters generate heterogeneity in the velocity (not shown) and temperature fields and (c) dilute regions of particles are heated rapidly, while denser clusters of cold particles persist further into the domain. Images correspond to a instantaneous snapshots for $\langle\varepsilon_p\rangle = 0.001$, Pe $= 5$ and $\chi = 829$. A high-resolution video of this image can be found in the supplementary materials.

where $\theta_{f/p}$ is the nondimensional temperature (for the fluid or particle phase) given by

$$\theta_{f/p} = \frac{T_{f/p} - T_{p,0}}{T_{f,0} - T_{p,0}} \tag{5.3}$$

and $\hat{x} = d/d_p$ is the nondimensional streamwise position.

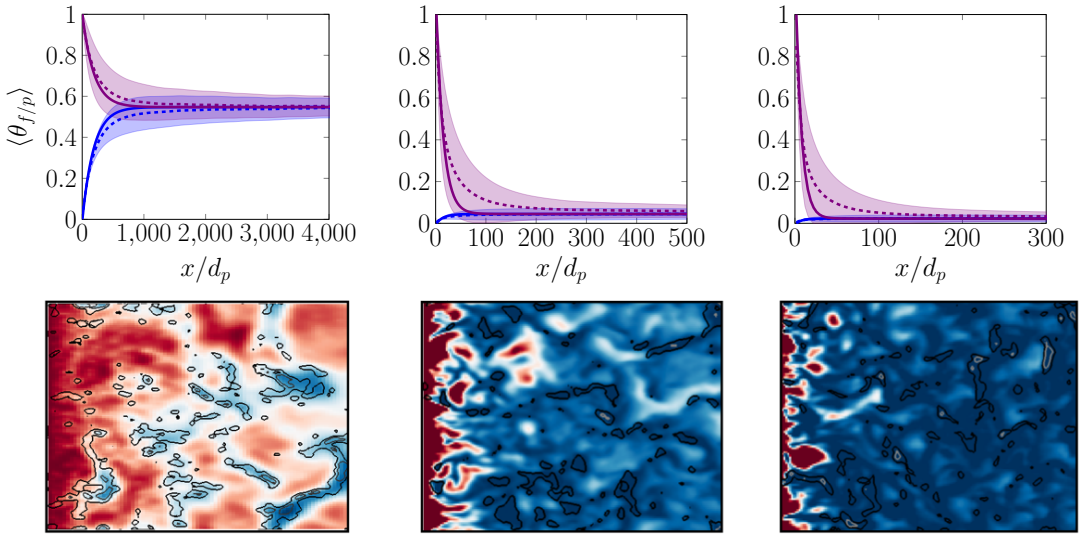


Figure 5.4: Temperatures are compared for the three volume fractions under study ($\langle\varepsilon_p\rangle = 0.001, 0.0255$ and 0.05, from left to right) and $(\mathrm{Pe}, \chi) = (5, 829)$. The top row shows the mean temperature profiles for a uniform distribution of particles (—, —) and the Euler–Lagrange simulations (- -, - -), where the shaded regions represent the $\pm\sigma$, where σ is the standard deviation. The bottom row shows the fluid temperature in the region between the inlet and $\hat{x} = l_{th}$. Red corresponds to high temperature and blue to low. The contours denote $\varepsilon_p = 2\langle\varepsilon_p\rangle$.

The thermal entrance lengths for the clustered, Euler–Lagrange results are compared with the development lengths for a uniform distribution of particles of equivalent mean volume fraction, l^0_{th} (see Figs. 5.4 and 5.5). By making this comparison, the effect of heterogeneity on interphase heat exchange can be isolated. Further, since the effect of clustering appears as a subgrid scale term in coarse-grained models [191, 56], the ratio of these quantities highlights the errors associated with neglecting these contributions.

For the cases considered, it is observed that the thermal equilibrium temperature is lowered with increasing volume fraction, owing to the increased mass loading of cool

particles as previously discussed. Additionally, the thermal entrance length decreases with increasing volume fraction, but in all cases, the presence of clusters acts to increase the thermal entrance length as compared with an uncorrelated distribution of particles. This can be seen in greater detail in Figure 5.3 and is primarily a consequence of the reduced contact with a hot fluid phase, making clustered particles less effective at cooling the surrounding fluid than lone particles. Finally, in Figure 5.4, the shaded regions represent ± 3 standard deviations from the mean temperature. This variation in temperature is greater in the fluid phase as compared with the particle phase, and the overall variation in temperature reduces with increasing volume fraction.

In Figure 5.5, the entrance length obtained from the simulations are normalized by l_{th}^0, as previously mentioned, and compared against volume fraction, Pe and χ. Here, we observe that for all the configurations considered, the entrance length for clustered particles is between 2 and 3 times longer than a uniform distribution, but that this relationship is complexly related to volume fraction and Péclet number, in particular. Notably, the development length increases *non-monotonically* with particle volume fraction for moderate and high Péclet numbers, which is likely explained by similar behavior observed in the normalized standard deviation of the volume fraction, $\sqrt{\langle \varepsilon_p'^2 \rangle}/\langle \varepsilon_p \rangle$, a measure of the degree of clustering [see 59]. Finally, we observe that the ratio of heat capacities, χ, has a relatively minimal effect on the thermal entrance length for clustered flows as compared with the entrance length for unclustered flows, l_{th}^0. This is shown in the inset panels in Figure 5.5 and indicates that the thermal entrance length for clustered flows does not change significantly as compared l_{th}^0. This implies that models for capturing heterogeneous behavior should depend only on $\langle \varepsilon_p \rangle$ and the Péclet number.

Finally, we compute l_{th}^0 over a range of volume fraction, Péclet , Reynolds and Prandtl numbers, and identify the following scaling relation for the thermal entrance

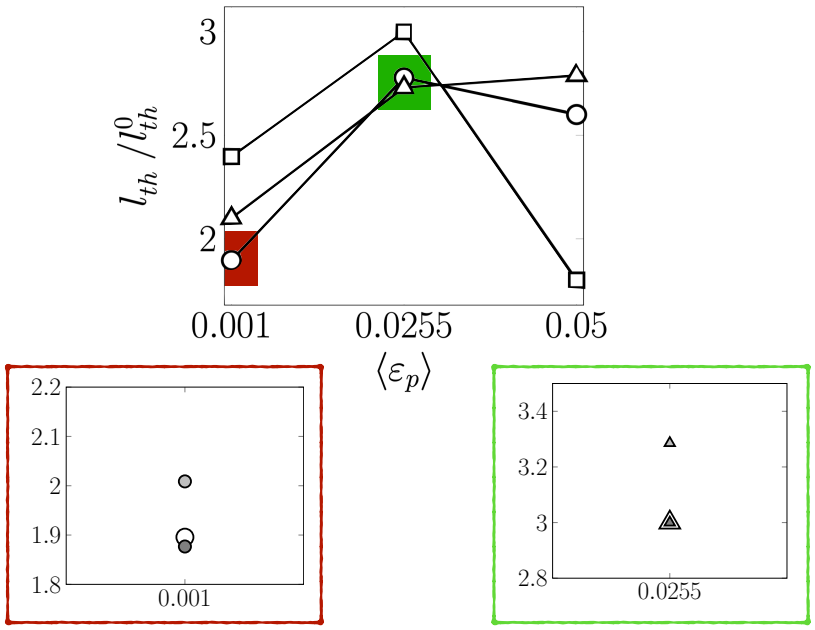


Figure 5.5: The entrance length normalized by the entrance length for a uniform distribution of particles of equivalent volume fraction (top). Here, Péclet numbers 1, 5 and 7 are denoted by squares, circles and triangles, respectively. The inset bottom two plots examine the effect of χ, where white, light gray and dark gray denote $\chi = (829, 909, 2270)$.

length for a uniform distributions of particles,

$$l^0_{th} = 0.108 \, \mathrm{Re}_{\mathrm{bulk}} \, \mathrm{Pr} \, \langle \varepsilon_p \rangle^{-1}. \tag{5.4}$$

In this expression, the existence of the particles augments the single-phase expression (5.1) by a factor of $0.216\langle \varepsilon_p \rangle^{-1}$. This quantifies the observation that the entrance length increases with decreasing solids volume fraction and increasing Reynolds number. Here, this increase is nearly exponential with respect to volume fraction and linear with respect to $\mathrm{Re}_{\mathrm{bulk}}$. The l-2 norm of the error of the scaling relation for $\mathrm{Re}_{\mathrm{bulk}} \in [0.2, 22]$ and $\langle \varepsilon_p \rangle \in [0.001, 0.5]$ is 0.02. A similar scaling relation can be

formulated for clustered flows, given as

$$l_{th} = 0.64 \frac{\sqrt{\langle \varepsilon_p'^2 \rangle}}{\langle \varepsilon_p \rangle} \left(0.1 \frac{\mathrm{Re}_{\mathrm{bulk}}}{\langle \varepsilon_p \rangle} + 0.02 \, \mathrm{Re}_{\mathrm{bulk}}^3 \right) + 0.108 \, \mathrm{Re}_{\mathrm{bulk}} \, \mathrm{Pr} \, \langle \varepsilon_p \rangle^{-1}, \tag{5.5}$$

where the variance in volume fraction is informed by a modified version of the model developed by Issangya et al. [198] given as

$$\sqrt{\langle \varepsilon_p'^2 \rangle} = 1.48 \langle \varepsilon_p \rangle \left(0.55 - \langle \varepsilon_p \rangle \right). \tag{5.6}$$

In this expression, the first coefficient differs from the original model of Issangya et al. [198], 1.584, to fit our data more accurately. The model (5.5) returns the scaling for an unclustered configuration, l_{th}^0, when particles are uncorrelated (i.e., $\langle \varepsilon_p'^2 \rangle = 0$). This scaling relation has a normalized l-2 error norm of 0.04 for the data considered in this study.

In the following section, we quantify the terms responsible for the complex behavior we observed in the thermal entrance length and propose closure to predict it over a range of conditions.

5.5 Modeling

In the previous section, we demonstrated that the thermal entrance length for clustered flows varies significantly from their uniform flow counterparts. Additionally, we observed that these differences depend complexly on the mean particle volume fraction as well as the Péclet number. To quantify the effect that correlated phases has on this phenomenon, we first derive the one-dimensional, two-fluid heat equations. Next, we evaluate the contributions of each of the terms appearing in the thermal balance and propose models for the dominant unclosed term.

5.5.1 One-dimensional heat equations

For the configurations under consideration, the flow is statistically stationary in time, statistically homogeneous in the spanwise directions and thermally evolving in the streamwise direction. This implies that all quantities of interest are one-dimensional in x. To formulate the associated 1D heat equations, we first nondimensionalize the heat equation, then conduct volume fraction-weighted (phase) averaging. Nondimensionalization is carried out by selecting the particle diameter, d_p, as a characteristic length scale and the inlet bulk velocity, u_{bulk}, as a characteristic velocity. Details on both of these derivations can be found in A and B.

Beginning with the fluid-phase heat equation (2.17), the nondimensional fluid temperature equation is given by

$$\text{Pe}\frac{\partial}{\partial \hat{t}}\left(\varepsilon_f \theta_f\right) + \text{Pe}\frac{\partial}{\partial \hat{\boldsymbol{x}}}\left(\varepsilon_f \hat{\boldsymbol{u}}_f \theta_f\right) = \varepsilon_f \frac{\partial^2 \theta_f}{\partial \hat{\boldsymbol{x}}^2} - 6\text{Nu}\varepsilon_p\left(\theta_f - \theta_p\right), \tag{5.7}$$

where $\hat{x} = x/d_p$, $\hat{u}_f = u_f/u_{\text{bulk}}$ and $\hat{t} = t/(d_p/u_{\text{bulk}})$. The particle-phase heat equation (2.25) can be similarly nondimensionalized. First, the particle phase heat equation is rewritten in the Eulerian sense by conducting a change of frame from the Lagrangian particle heat equation and projecting it to the Eulerian grid (see A). Then, in the same manner as the fluid heat equation, nondimensionalization yields

$$\chi\ \text{Pe}\frac{\rho_p}{\rho_f}\left[\frac{\partial\left(\varepsilon_p \theta_p\right)}{\partial \hat{t}} + \frac{\partial\left(\varepsilon_p \hat{\boldsymbol{u}}_p \theta_p\right)}{\partial \hat{\boldsymbol{x}}}\right] = \varepsilon_p \frac{\partial^2 \theta_f}{\partial \hat{\boldsymbol{x}}^2} + 6\text{Nu}\varepsilon_p\left(\theta_f - \theta_p\right), \tag{5.8}$$

where, χ is the ratio of heat capacities.

Next, Reynolds averages (denoted with angled brackets) in time and the spanwise directions are applied in order to treat these expressions as statistically one-dimensional. In doing so, the time derivative is null and due to periodicity in the $y-$ and $x-$ directions, gradients and divergence operators reduce to full derivatives with

respect to x. This yields

$$\mathrm{Pe}\frac{\mathrm{d}\langle\varepsilon_f\hat{u}_f\theta_f\rangle}{\mathrm{d}\hat{x}} = \frac{\mathrm{d}^2\langle\varepsilon_f\theta_f\rangle}{\mathrm{d}\hat{x}^2} - \langle 6\mathrm{Nu}\varepsilon_p\left(\theta_f-\theta_p\right)\rangle \tag{5.9}$$

and

$$\chi\,\mathrm{Pe}\frac{\rho_p}{\rho_f}\frac{\mathrm{d}\langle\varepsilon_p\hat{u}_p\theta_p\rangle}{\mathrm{d}\hat{x}} = \frac{\mathrm{d}^2\langle\varepsilon_p\theta_p\rangle}{\mathrm{d}\hat{x}^2} + \langle 6\mathrm{Nu}\varepsilon_p\left(\theta_f-\theta_p\right)\rangle. \tag{5.10}$$

As will be discussed later, the diffusion terms are found to have a minimal contribution to the thermal balance in both phases, but they are included here since the Péclet numbers are $\mathcal{O}(1)$. In cases of very large Péclet number, however, the diffusion term can be eliminated *a priori* due to the factor $1/\mathrm{Pe}$ that multiplies it.

Due to the presence of the volume fraction on all terms, a phase average defined as $\langle(\cdot)\rangle_f = \langle\varepsilon_f(\cdot)\rangle/\langle\varepsilon_f\rangle$ and $\langle(\cdot)\rangle_p = \langle\varepsilon_p(\cdot)\rangle/\langle\varepsilon_p\rangle$ (as described in Fox [39]) is convenient to invoke. This process substantially reduces the number of terms present as compared with strict Reynolds averaging.

In these expressions, angled brackets without a subscript, $\langle(\cdot)\rangle$, denote a Reynolds average in time and the cross stream directions (i.e., y and z). Here, fluctuations from Reynolds averages are denoted as $(\cdot)'$ and fluctuations from the particle and fluid phase averages are denoted as $(\cdot)''$ and $(\cdot)'''$, respectively. This yields a coupled system of two-fluid equations that may be utilized to model macroscopic heat transfer (noting that the solution of the momentum equations is trivial for the configuration under consideration), given as

$$\begin{aligned}\langle\hat{u}_f\rangle_f\frac{\mathrm{d}\langle\theta_f\rangle_f}{\mathrm{d}\hat{x}} - \frac{1}{\mathrm{Pe}}\frac{\mathrm{d}^2\langle\theta_f\rangle_f}{\mathrm{d}\hat{x}^2} &= \underbrace{-\frac{\mathrm{d}}{\mathrm{d}\hat{x}}\langle\hat{u}_f'''\theta_f'''\rangle_f}_{\text{Term 1}} \\ -\frac{6\langle\varepsilon_p\rangle}{\mathrm{Pe}\,\langle\varepsilon_f\rangle}\Big[&\underbrace{\langle\mathrm{Nu}\rangle_p\left(\langle\theta_f\rangle_f-\langle\theta_p\rangle_p\right)}_{\text{Term 2}} + \underbrace{\langle\mathrm{Nu}\rangle_p\langle\theta_f'''\rangle_p}_{\text{Term 3}} + \underbrace{\langle\mathrm{Nu}''\theta_f''\rangle_p}_{\text{Term 4}} - \underbrace{\langle\mathrm{Nu}''\theta_p''\rangle_p}_{\text{Term 5}}\Big]\end{aligned} \tag{5.11}$$

and

$$\langle \hat{u}_p \rangle_p \frac{\mathrm{d}\langle \theta_p \rangle_p}{\mathrm{d}\hat{x}} - \frac{\rho_f}{\rho_p \chi \, \mathrm{Pe}} \frac{\mathrm{d}^2 \langle \theta_p \rangle_p}{\mathrm{d}\hat{x}^2} = - \underbrace{\frac{\mathrm{d}}{\mathrm{d}\hat{x}} \langle \hat{u}_p'' \theta_p'' \rangle_p}_{\text{Term 6}}$$

$$+ \frac{6\rho_f}{\rho_p \chi \, \mathrm{Pe}} \Big[\underbrace{\langle \mathrm{Nu} \rangle_p \left(\langle \theta_f \rangle_f - \langle \theta_p \rangle_p \right)}_{\text{Term 2}} + \underbrace{\langle \mathrm{Nu} \rangle_p \langle \theta_f''' \rangle_p}_{\text{Term 3}} + \underbrace{\langle Nu'' \theta_f'' \rangle_p}_{\text{Term 4}} - \underbrace{\langle \mathrm{Nu}'' \theta_p'' \rangle_p}_{\text{Term 5}} \Big]. \tag{5.12}$$

The terms in these expressions can be categorized as purely fluid, purely particle and mixed. The terms on the left-hand side of the particle and fluid equations represent convection and diffusion and are purely fluid and purely particle, respectively. Terms 1 and 6 are scalar fluxes, which are unclosed and traditionally modelled by classical gradient diffusion models (e.g., the Boussinesq approximation). While these methods are successful in single-phase flows, they have been shown to fall short of being predictive in the context of highly anisotropic, multi-phase flows [164]. Finally, the interphase heat exchange terms (Terms 1–5), are the same across the fluid and particle phase descriptions, with the exception of a constant factor of $\rho_f/(\rho_p\chi)$ that appears in the particle phase equation and a factor of $\langle \varepsilon_p \rangle / \langle \varepsilon_f \rangle$ in the fluid phase. For brevity, these two factors are referred to as C_1 and C_2 henceforth. Of the interphase heat exchange terms, only Term 2 is a function of solution variables ($\langle \theta_f \rangle_f, \langle \theta_p \rangle_p$) and are therefore closed. Term 3 includes a covariance between volume fraction and fluid temperature fluctuations, $\langle \theta_f''' \rangle_p$, which has been shown to be the main contributor to hindering heat transfer in temporally evolving, homogeneous systems [56]. Terms 4 and 5 are cross correlations between phase temperature and Nusselt number.

In the absence of clustering (e.g., no correlation between temperature and volume fraction), the only terms that remain are convection, diffusion and Term 2. Following from the definition of the phase average, $\langle \theta_f \rangle_f = \langle \theta_f \rangle + \langle \varepsilon_f' \theta_f' \rangle / \langle \varepsilon_f \rangle$, thus, in an uncorrelated, homogeneous system, $\langle \theta_f \rangle_f$ is equivalent to $\langle \theta_f \rangle$ and $\langle \theta_p \rangle_p = \langle \theta_p \rangle$) since cross correlations are null.

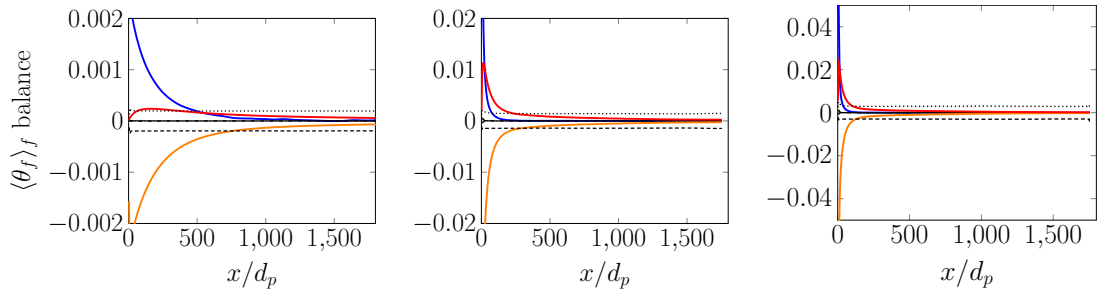


Figure 5.6: Balance of terms contributing to the phase averaged fluid temperature as given in Eq. (5.11) for $Pe = 5$, $\chi = 829$. In each of the three volume fractions (0.001, 0.0255 and 0.05 from left to right), three dominate the thermal behavior: Convection (blue), Term 2 (orange) and Term 3 (red).

5.5.2 Thermal budget

To guide our modeling efforts, we now evaluate which of the terms appearing in (5.11) and (5.12) have leading order effects. As such, the balance of these terms is shown in Figure 5.6 for the illustrative case of Pe = 5 and $\langle\varepsilon_p\rangle = 0.0255$. This demonstrates that for the configuration under consideration, thermal behavior is dominated by convection, Term 2 and Term 3. Of these terms, only the fluid phase temperature fluctuations as seen by the particles, $\langle\theta_f'''\rangle_p$ (defined as the 'drift temperature' in Rauchenzauner and Schneiderbauer [191]), requires modeling and is equivalent across both phases.

5.5.3 Closure of the drift temperature

In this section, we propose a closure model for $\langle\theta_f'''\rangle_p$ and equivalently, $\langle\varepsilon_p'\theta_f'\rangle$ (see C). As previously mentioned, and by definition of the phase average, the phase averaged temperatures are comprised of the Reynolds averaged temperature plus the cross correlation between volume fraction and temperature (i.e., $\langle\theta_f\rangle_f = \langle\theta_f\rangle + \langle\varepsilon_f'\theta_f'\rangle/\langle\varepsilon_f\rangle$). Because of this, specifying boundary conditions for the heat equations in terms of phase-averaged quantities cannot be done *a priori* without an additional closure for these contributions. Rather than providing additional closures

(one each for $\langle \varepsilon_p' \theta_p' \rangle$ and $\langle \varepsilon_f' \theta_f' \rangle$), we note that for the configuration under study, the cross correlations are constant with respect to the streamwise direction and only shift the temperature solution by this amount. In other words, the thermal entrance length is equivalent when considering either the phase averaged *or* Reynolds averaged temperatures, but the Reynolds averaged formulation does not require special treatment for the boundary conditions, as they are specified as the same for clustered and unclustered flows. (See C for more details).

Due to the equivalency of $\langle \varepsilon_p' \theta_f' \rangle$ and $\langle \theta_f''' \rangle_p$, the model proposed herein is suitable for use in simulations for which the solution variables are phase-averaged *or* Reynolds averaged (demonstrated in Figure 5.8 where the proposed model detailed in Section 6.4 is used in forward solutions of both sets of state variables). Thus implying it is appropriate for use in a general two-fluid solver in which the hydrodynamics and thermodynamics evolve simultaneously. Of course, in this situation, additional closures are required for the fluid and particle momentum equations in order to capture cross correlations.

We begin from the simplified, Reynolds averaged equations,

$$\langle \hat{u}_f \rangle \frac{\mathrm{d}\langle \theta_f \rangle}{\mathrm{d}\hat{x}} = -\frac{6\langle \varepsilon_p \rangle \widetilde{Nu}}{\mathrm{Pe}\langle \varepsilon_f \rangle} \left[\langle \theta_f \rangle - \langle \theta_p \rangle + \frac{\langle \varepsilon_p' \theta_f' \rangle}{\langle \varepsilon_p \rangle} \right], \tag{5.13}$$

where $\widetilde{Nu}$ denotes the Nusslet number computed using the correlation proposed by Sun and Zhu [171] and mean quantities as arguments. As detailed in C, all of the unclosed Reynolds averaged terms are null, except for the cross correlation between particle volume fraction and the fluid-phase temperature fluctuations arising from Term 3, as was observed for temporally evolving gas-particle flows from Guo and Capecelatro [56].

This result points to the fact that cross-correlations between volume fraction and temperature shift the phase averaged temperature from the Reynolds averaged tem-

perature (e.g., $\langle\theta_f(\hat{x})\rangle_f = \langle\theta_f(\hat{x})\rangle + \langle\varepsilon'_f\theta'_f\rangle$), where in these configurations the cross correlations are constant with respect to $\hat{x}$.

In formulating the closure for $\langle\varepsilon'_p\theta'_f\rangle$, we observe that all configurations considered in this work satisfy the following scaling relation

$$\frac{1}{(\langle\theta_f\rangle - \langle\theta_p\rangle)}\left(\frac{\mathrm{d}\langle\theta_f\rangle}{\mathrm{d}\hat{x}} + C_1\left(\langle\theta_f\rangle - \langle\theta_p\rangle\right)\right) = b\left(\langle\theta_f\rangle - \langle\theta_p\rangle + 1\right) \tag{5.14}$$

where b is a constant coefficient, which may depend upon $\langle\varepsilon_p\rangle$, χ and Pe. Owing to this relation, we impose that the the proposed model is of the form

$$\frac{\langle\varepsilon'_p\theta'_f\rangle}{\langle\varepsilon_p\rangle} = -\frac{b}{C_1}\left(\langle\theta_f\rangle - \langle\theta_p\rangle\right)\left(\langle\theta_f\rangle - \langle\theta_p\rangle + 1\right)) \tag{5.15}$$

and the system of equations is given as

$$\frac{\mathrm{d}\langle\theta_f\rangle}{\mathrm{d}\hat{x}} = -C_1\left(\langle\theta_f\rangle - \langle\theta_p\rangle\right) + b\left(\langle\theta_f\rangle - \langle\theta_p\rangle\right)\left[\left(\langle\theta_f\rangle - \langle\theta_p\rangle\right) + 1\right] \tag{5.16}$$

$$\frac{\mathrm{d}\langle\theta_p\rangle}{\mathrm{d}\hat{x}} = C_2\left(\langle\theta_f\rangle - \langle\theta_p\rangle\right) - \frac{C_2}{C_1}b\left(\langle\theta_f\rangle - \langle\theta_p\rangle\right)\left[\left(\langle\theta_f\rangle - \langle\theta_p\rangle\right) + 1\right]. \tag{5.17}$$

The open-source, Gene Expression Programming (GEP) MATLAB code of Searson [199], is leveraged to learn the dependence of b on operating parameters. The resultant model was selected from the models learned using a population size of 300, with 500 generations and a maximum number of genes per individual of 6. The GEP algorithm was provided with the value of b and associated $\langle\varepsilon_p\rangle$ and Pe for each training case (all three volume fractions, all three Péclet numbers and $\chi = 829$) and was permitted to evolve expressions from the following mathematical operations: multiplication/division, addition/subtraction, exponential/log, and square/cube. The resultant learned model for b is given as

$$b = \left(1.16\ln(\langle\varepsilon_p\rangle) - 0.335\mathrm{Pe} + 5.85\langle\varepsilon_p\rangle\mathrm{Pe} + 19.7\right)\sqrt{\langle\varepsilon'^2\rangle}\left(1 - e^{-\langle\varepsilon_p\rangle/\mathrm{Pe}}\right), \tag{5.18}$$

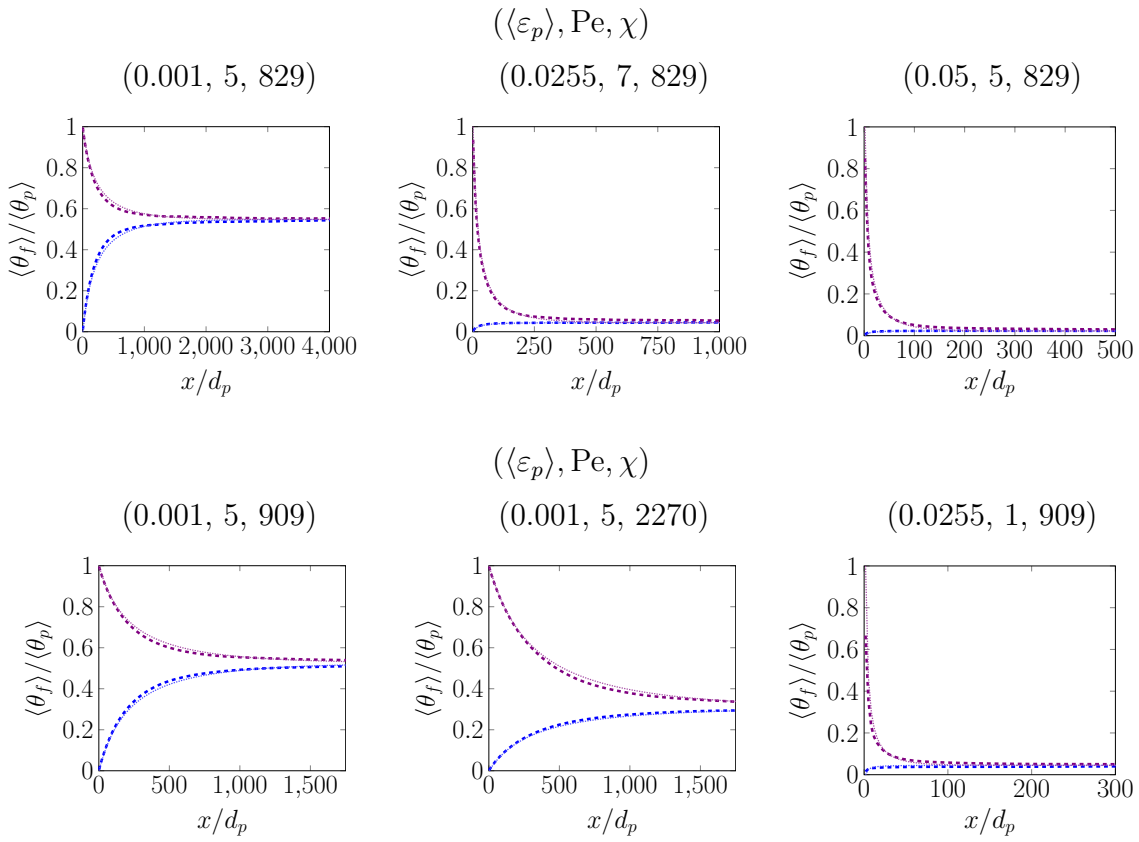


Figure 5.7: Three example instances each of model performance (,) on training data (top row) and testing data (bottom row), as compared to the mean quantities from Eulerian–Lagrangian data (- -, - -).

where the inclusion of the variance of volume fraction, $\sqrt{\langle\varepsilon'^2\rangle}$, in the expression for b ensures proper asymptotic behavior in the limit of no clustering (i.e., Term 3 vanishes in the case of a uniform distribution of particles), which is modeled according to (5.6).

Figure 5.7 highlights the forward solution of the proposed model and the forward solution for a uniform distribution of particles. Both are compared against the mean Euler–Lagrange data. The top and bottom rows show three representative training and test cases, respectively. Since the ratio of heat capacities was observed to have a minimal effect on entrance length as discussed in Section 5.4, perturbations in χ were reserved for the testing set.

Additionally, the predicted entrance lengths for the uncorrelated forward solution and the forward solution with the proposed model for b is summarized in Table 5.3 and

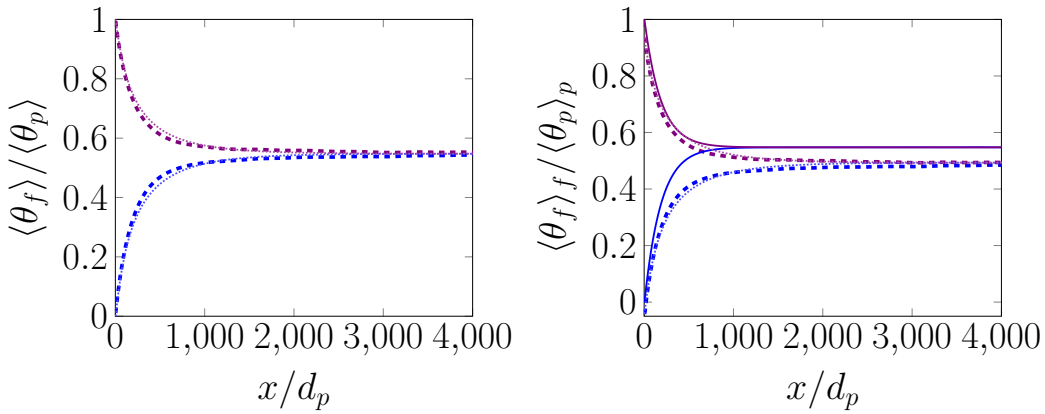


Figure 5.8: The learned model (shown as (·····, ·····) and described in Eq. (5.18)) demonstrates improved prediction of thermal entrance length as compared to the Euler–Lagrange results (- - , - -) in both the Reynolds-averaged (left) and the phase-averaged formulations (right). The forward solution using the assumption of uniformly distributed particles is shown as (—, —), and is the same in both plots since phase averaging and Reynolds averaging are equivalent when the phases are uncorrelated.

compared against the Euler–Lagrange results. We find that using an assumption of uniformly distributed particles results in an under prediction of the thermal entrance length between 40 and 70%. This highlights the importance of incorporating local particle heterogeneity in reduced-order (coarse grained) models. The proposed model demonstrates improved performance, predicting the thermal entrance length within 3.6%, on average for the training dataset and within 8.6% for the testing dataset.

To make physical connections with the resultant model, it is helpful to introduce a new variable representing the temperature difference between the phases, $\langle\theta_\Delta\rangle = \langle\theta_f\rangle - \langle\theta_p\rangle$, and corresponding transport equation. This definition and some algebraic

χ	Pe	$\langle\varepsilon_p\rangle$	l_{th}	l_{th}^0	l_{th}^{model}	ϵ^0	ϵ^{model}
840	1	0.001	258.9	108.0	214.3	0.6	0.026
		0.0255	36.0	12.0	32.5	0.7	0.095
		0.005	15.4	8.6	15.4	0.4	0.00
	5	0.001	1,059.5	558.9	1,114.4	0.5	0.045
		0.0255	128.6	46.3	152.5	0.6	0.053
		0.005	66.9	25.7	66.9	0.6	0.000
	7	0.001	1,681.8	780.1	1,555.0	0.5	0.084
		0.0255	173.2	63.4	210.9	0.6	0.020
		0.05	90.9	32.6	94.3	0.6	0.000
Testing Dataset							
921	1	0.0255	39.4	1.2	32.5	0.7	0.174
	5	0.001	1,177.8	584.6	1,164.1	0.5	0.095
2300	1	0.0255	36.0	1.2	32.5	0.7	0.014
	5	0.001	1,582.4	848.6	1,695.5	0.5	0.060

Table 5.3: Summary of thermal entrance lengths normalized by d_p for clustered gas–solid flows and associated model errors. The learned model was trained on data for $\chi = 840$. Remaining cases were reserved for testing. On average, the entrance length predicted using an uncorrelated particle-phase assumption is under predictive by 58%, while the prediction from the learned model predicts entrance length within 5.1%, where the mean training and testing errors are 3.6% and 8.6%, respectively.

manipulation, yields the following system:

$$\frac{\mathrm{d}\langle\theta_f\rangle}{\mathrm{d}\hat{x}} = (b - C_1)\langle\theta_\Delta\rangle\left(1 - \frac{\langle\theta_\Delta\rangle}{(b - C_1)/b}\right) \tag{5.19}$$

$$\frac{\mathrm{d}\langle\theta_p\rangle}{\mathrm{d}\hat{x}} = \frac{C_2(C_1 - b)}{C_1}\langle\theta_\Delta\rangle\left(1 - \frac{\langle\theta_\Delta\rangle}{(b - C_1)/b}\right) \tag{5.20}$$

$$\frac{\mathrm{d}\langle\theta_\Delta\rangle}{\mathrm{d}\hat{x}} = \frac{\mathrm{d}\langle\theta_f\rangle}{\mathrm{d}\hat{x}} - \frac{\mathrm{d}\langle\theta_p\rangle}{\mathrm{d}\hat{x}} \tag{5.21}$$

$$= \left(-(C_1 + C_2) + \frac{b(C_1 - C_2)}{C_1}\right)\langle\theta_\Delta\rangle\left(1 - \frac{\langle\theta_\Delta\rangle}{(b - C_1)/b}\right), \tag{5.22}$$

where we note that the equation for the mean temperature difference is of the same form as the logistic equation, i.e., $\mathrm{d}A/\mathrm{d}x = kA(1-A/L)$. In this sense, L is frequently referred to as the limiting factor, or carrying capacity, of the system and k is the growth rate. In the context of heat transfer for particle-laden flows, bifurcation points

exist when either $\langle\theta_\Delta\rangle = 0$ or $\langle\theta_\Delta\rangle = L$. For this system and boundary conditions, the only physically relevant bifurcation occurs when the temperature difference is null. This point is a stable attractor, ensuring that all realizations with physical boundary conditions and parameters will relax to thermal equilibrium.

Finally, the growth rate (which in this case is a negative value, indicating decay to equilibrium) is given as $-(C_1+C_2)+b(C_1-C_2)/C_1$. In the event of no clustering, the rate reduces to the uncorrelated growth rate $-(C_1+C_2)$, thus demonstrating that the presence of clusters impedes the rate at which the phases approach equilibrium. Written in this form, it can also be observed that the fluid- and particle-phase growth rates differ by a factor of $(-C_1/C_2)$, when clustering is present. Further, due to the dependence of b on volume fraction and Péclet number, the model quantifies the complex interplay of volume fraction and Péclet number on thermal entrance length. This effect is visualized in Figure 5.9, where we observe that for low Péclet numbers, variations in volume fraction have a greater effect on the value of b. Similarly, at high volume fraction, changes in Péclet number (particularly between 0 and 1) also result in large changes in b. Conversely, at high Péclet number and as volume fraction approaches null, b only changes slightly. Further, since b implicates heat transfer impedance, one can expect longer thermal entrance lengths for lower volume fractions at higher Péclet numbers and lower particle volume fractions.

5.6 Conclusions

In this work, high resolution Euler–Lagrange simulations were leveraged to understand the effect of heterogeneity on the thermal entrance length. These computations enabled the quantification of the complex dependency of the entrance length on relevant simulation parameters, Péclet number, volume fraction and ratio of phase heat capacities. In addition, we compared the thermal entrance length for clustered and uniform distributions of particles and found that clustering causes a 2 to 3 fold in-

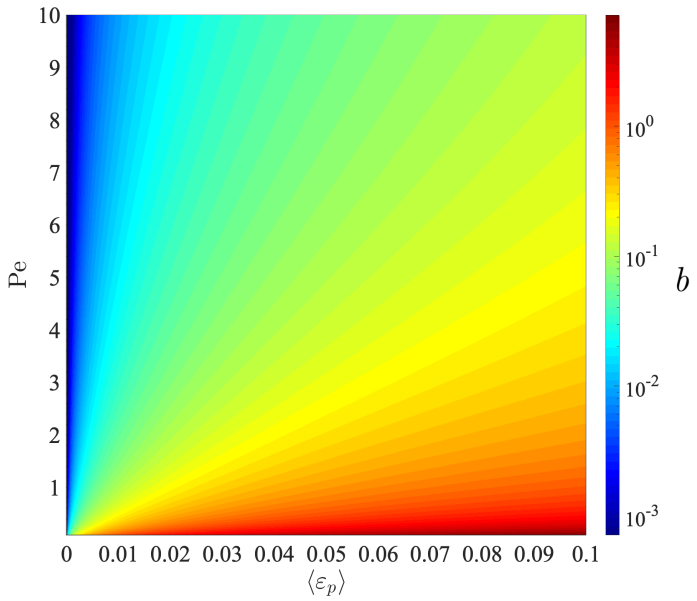


Figure 5.9: The modeled quantity, b (5.18), shown with respect to $\langle\varepsilon_p\rangle$ and Pe.

crease in l_{th}. To capture this effect, we propose a scaling relation for l_{th} in Eq. (5.4) (for uniform distribution of particles) and Eq. (5.5) (for clustered) that bares resemblance to scaling laws for the thermal entrance length of single-phase flows, but with an additional factor to account for the presence of particles.

To identify the physics responsible for the change in thermal behavior of clustered flows, we derive the 1D two-fluid heat equations and evaluate which terms dominate. This analysis demonstrated that the delay in heat transfer is described entirely by the covariance between volume fraction and fluid temperature fluctuations, also known as the 'drift temperature.' Since this quantity is sensitive to variations in Péclet number and mean particle volume fraction, but is minimally sensitive to the ratio of heat capacities, we then leverage scaling arguments and Gene Expression Programming to propose a closure. The resultant model captures the complex dependency of the drift temperature on Pe and $\langle\varepsilon_p\rangle$ and reduces the error in predicting thermal entrance length by 90% as compared to predictions that neglect heterogeneity. We also note that the proposed model is appropriate for use in both the Reynolds aver-

aged and phase averaged formulations of the heat equations, making it a suitable for use in Euler–Euler codes in which the thermodynamics and hydrodynamics evolve simultaneously.

CHAPTER VI

Toward More Efficient Modeling: Blending Sparse Regression with Gene Expression Programming

In recent years, there has been an explosion of machine learning techniques for turbulence closure modeling, though many rely on augmenting existing models. While this has proven successful in single-phase flows, it breaks down for multiphase flows, particularly when the system dynamics are controlled by two-way coupling between the phases. In this work, we propose an approach that blends sparse regression and gene expression programming (GEP) to generate closed-form algebraic models from simulation data. Sparse regression is used to determine a minimum set of functional groups required to capture the physics and GEP is used to automate the formulation of the coefficients and dependencies on operating conditions. The framework is demonstrated on a canonical gas–solid flow in which two-way coupling generates and sustains fluid-phase turbulence.

6.1 Introduction

In the last decade, data-driven approaches have become the predominant tool for developing turbulence models [21]. Of these approaches, Neural Networks (NNs) have gained a considerable amount of traction [22, 23, 24, 25, 26, 21, 27, 28, 29, 30].

In contrast, a relative minority of approaches have elected to pursue strategies that enable a compact, algebraic closure. Formulating models in this way has several important properties including increased interpretability, ease of dissemination and straightforward integration into existing solvers. These techniques generally fall into two categories, (1) symbolic regression and (2) gene expression programming.

In the case of sparse regression, Brunton et al. [38] developed a strategy based on sparse regression that identifies the underlying functional form of the nonlinear physics by optimizing a coefficient matrix that acts upon a matrix of trial functions. This construct has the important benefit of including the user in the modeling process through selection of the trial functions. Schmelzer et al. [170], Beetham and Capecelatro [58] recently extended the sparse identification framework of Brunton et al. [38] to infer algebraic stress models for the closure of the Reynolds-averaged Navier–Stokes (RANS) equations. In Schmelzer et al. [170], the models are written as tensor polynomials and built from a library of candidate functions. In Beetham and Capecelatro [58], Galilean invariance of the resulting models are guaranteed through thoughtful tailoring of the feature space.

Gene expression programming (GEP), a data-driven technique inspired by the Darwinian concept of survival-of-the-fittest, heuristically evolves symbolic models until error is reduced beyond a threshold. In recent years, this strategy has gained attention in the turbulence modeling. For example, GEP has demonstrated success in the contexts of modeling large eddy simulation (LES) subgrid scale closures [200], boundary layer theory [201], turbulent pipe flow [202] and informing RANS closures [203, 204].

While these data-driven techniques have been increasingly utilized for modeling single-phase turbulence, their application to multiphase turbulence modeling is still relatively uncharted. Despite this, multiphase flows present a rich and diverse class of problems for which machine learning can prove useful. Due to the large param-

eter space frequently attributed to such flows, traditional modeling techniques have historically failed, especially beyond dilute regimes, where models extended from single-phase turbulence break down [see, e.g., 39, and discussion therein]. This divergence from single-phase turbulence theory can be attributed to the fact that at moderate volume fractions, particles generate turbulent kinetic energy (TKE) at the smallest scales. This is the direct antithesis to the classical notion of the turbulent energy cascade. Additionally, numerous practical applications span regimes from dilute to dense particle loadings, motivating the need for models that are accurate across regimes. This motivates the need for methodologies capable of formulating closures 'from scratch,' rather than augmenting existing models. These challenges, along with the societal importance and pervasiveness of these flows, make them excellent candidates for improvements in data-driven modeling techniques.

In this work, we propose a blending modeling approach which combines the strengths of both sparse regression and GEP to inform multiphase turbulence closures in a way that leverages the physical knowledge of the modeler and automates the determination of model components for which physical insight is not obvious or does not exist. To demonstrate the utility of such an approach, we present a simple configuration in which strong two-way coupling between fluid and particles generates and sustains turbulence. This configuration has been discussed extensively in prior work [see, e.g., 40, 41, 58, for more exhaustive details] and serves as a case study here.

6.2 Methodology

It is well established [17] that any tensor quantity, $\mathcal{D}_{ij}$, can be exactly described by an infinite sum given as

$$\mathcal{D}_{ij} = \sum_{n=1}^{\infty} \beta^{(n)} \mathcal{T}_{ij}^{(n)}, \tag{6.1}$$

where $\beta^{(n)}$ represents the n-th coefficient associated with a corresponding basis tensor, $\mathcal{T}_{ij}^{(n)}$. The coefficients may range in complexity from constants to nonlinear functions of the principal invariants of the tensor bases. For many configurations, this infinite sum can reduced to a finite sum by leveraging the Cayley–Hamilton theorem. This results in a reduced set of tensors termed a minimal invariant basis [e.g., 16]. Using knowledge of the system physics, a minimal invariant basis can be derived. Once this basis is established, modeling can be broken into two tasks: (1) *Which of the basis tensors are most important for capturing the physics at play?* and (2) *How do the coefficients depend on principal invariants or system parameters?*

Sparse regression has been shown to be successful at addressing the first task [58, 59] and works well for the second task when *constant* coefficients are sufficient. However, when the system has a complex and large parameter space, as is the case for multiphase turbulence, constant coefficients are no longer sufficient for capturing physics across scales. In this situation, sparse regression is not an efficient method for determining the form of the coefficients and requires the modeler to supply all potential test functions to the algorithm manually. While this has important benefits for embedding physics-based reasoning and properties into the resultant model (e.g., form invariance), it implies a tedious, 'guess-and-check' exercise if physics-based arguments can no longer be used to supply test functions. In the present method, we propose to offload this work to a gene expression algorithm when naivity in functional form is unavoidable. This effectively automates the process of evaluating trial functions for the coefficients, while preserving the benefits of using sparse regression to inform the tensorial building blocks of the model.

The method can be summarized by three distinct modeling steps, as shown in Figure 6.1, and outlined here for data spanning s unique conditions in the parameter space (in the context of multiphase, turbulent flows, these parameters might include solids volume fraction, Reynolds number, etc.):

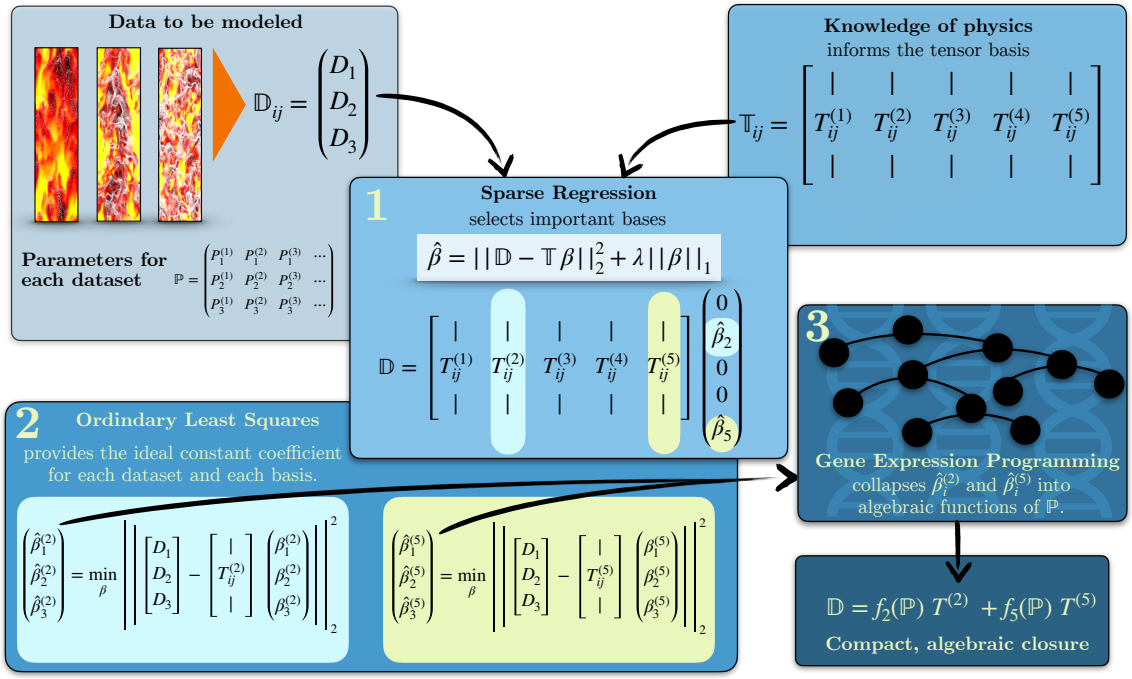


Figure 6.1: The modeling methodology has three steps: (1) Sparse regression identifies the important basis tensors, (2) OLS squares provides the ideal coefficients for each of the data sets for each of the identified bases and (3) GEP collapses the ideal coefficients for each case into a compact, algebraic closure.

1. Use sparse regression to identify the basis tensors required to describe physics by optimizing

$$\hat{\beta} = \min_{\beta} ||\mathbb{D} - \mathbb{T}\beta||_2^2 + \lambda||\beta||_1 \tag{6.2}$$

and assuming constant coefficients. Each base associated with a nonzero coefficient in $\hat{\beta}$ is deemed to be 'essential' and is retained in the final model. The surviving bases are then condensed into a subset of $\mathbb{T}$, denoted $T^{\subset}$.

2. For each of the s conditions, compute the ideal constant coefficients associated with the p essential bases, using Ordinary Least Squares (OLS):

$$\hat{\beta}_s^{\subset} = \min_{\hat{\beta}_s^{\subset}} ||D_s - T_s^{\subset}\beta_s^{\subset}||_2^2, \tag{6.3}$$

where $\hat{\beta}_s^{\subset}$ is of size $p \times 1$, D_s is size $q \times 1$, where q is the amount of data in the

dataset (e.g. the number of time steps) and $T_s^{\subset}$ is size $q \times p$. Note that both $T^{\subset}$ and D require tensorial data to be reorganized as vertical vectors [see 58, for details]. After this process has been done for all s conditions, concatenate each of the $\hat{\beta}_s^{\subset}$ vectors into a matrix of size $p \times s$.

3. Finally, provide each p-th row of $\hat{\beta}^{\subset}$ and matrix of parameters, $\mathbb{P}$, associated with the s conditions as input to the GEP algorithm. The resulting functional model for $\hat{\beta}^{\subset}$ effectively collapses the vector of discrete values for $\hat{\beta}^{\subset}$ to a continuous, closed form with algebraic dependence on system parameters.

This modeling flow is illustrated in Figure 6.1, where $s = 3$ and $p = 2$ for demonstration purposes.

6.3 Case study description

Multiphase flows span large parameter spaces, making modeling challenging. Thus, we use a simple gas–solid flow in which two-way coupling between the phases drives the underlying turbulence as a case study to evaluate the effectiveness of the proposed modeling framework. In this configurations, rigid spherical particles are suspended in an unbounded (triply periodic) domain containing an initially quiescent gas. As particles settle under the influence of gravity, they spontaneously form clusters. Due to momentum exchange between phases, particles entrain the fluid, generating turbulence therein. A frame of reference with the fluid phase is considered, such that the mean streamwise fluid velocity is null. Key non-dimensional numbers that characterize the system include the Reynolds number, the Archimedes number, defined as

$$\mathrm{Ar} = (\rho_p/\rho_f - 1) d_p^3 g/\nu_f^2. \tag{6.4}$$

Alternatively, a Froude number can be introduced to characterize the balance between gravitational and inertial forces, defined as $\mathrm{Fr} = \tau_p^2 g/d_p$, where $\tau_p = \rho_p d_p^2/(18\rho_f \nu_f)$ is

the particle response time.

The mean particle-phase volume fraction is varied from $0.001 \leq \langle \varepsilon_p \rangle \leq 0.05$ and gravity is varied from $0.8 \leq g \leq 8.0$. Here, angled brackets denote an average in all three spatial dimensions and time. Due to the large density ratios under consideration, the mean mass loading, $\varphi = \langle \varepsilon_p \rangle \rho_p / \left(\langle \varepsilon_f \rangle \rho_f \right)$, ranges from $\mathcal{O}(10)$–$\mathcal{O}(10^2)$, and consequently two-way coupling between the phases is important. A large enough domain with a sufficiently large number of particles is needed to observe clustering. To enable simulations on this scale, we use an Eulerian–Lagrangian approach [81]. Fluid equations are solved on a staggered grid with second-order spatial accuracy and advanced in time with second-order accuracy using the semi-implicit Crank–Nicolson scheme. Particles are tracked individually in a Lagrangian frame of reference and integrated using a second-order Runge–Kutta method. Particle data is projected to the Eulerian mesh using a Gaussian filter described in Capecelatro and Desjardins [81].

Derivation of the *single-phase* RANS equations is done by directly averaging the Navier–Stokes equations. Derivation of the *multiphase* RANS equations, however, will retain additional physics if averaging is performed on the volume-filtered Navier-Stokes equations [39]. Volume fraction weighted averages, or phase averaging (PA), analogous to Favre averaging of variable density flows, has previously been derived [41]. For the relatively simple configuration used here, which is homogeneous in all spatial directions, statistically stationary in time and symmetric in the counter-gravity direction, the transport equations for the fluid-phase Reynolds stresses can be reduced to two unique, non-zero components. In the streamwise direction this

equation is given as

$$\frac{1}{2}\frac{\partial\langle u_f'''^2\rangle_f}{\partial t} = \underbrace{\frac{1}{\rho_f}\left\langle p_f\frac{\partial u_f'''}{\partial x}\right\rangle}_{\text{pressure strain (PS)}} - \underbrace{\frac{1}{\rho_f}\left\langle \sigma_{f,1i}\frac{\partial u_f'''}{\partial x_i}\right\rangle}_{\text{viscous dissipation (VD)}} + \underbrace{\frac{\varphi}{\tau_p^\star}\left(\langle u_f''' u_p''\rangle_p - \langle u_f'''^2\rangle_p\right)}_{\text{drag exchange (DE)}} + \underbrace{\frac{\varphi}{\tau_p^\star}\langle u_f'''\rangle_p\langle u_p\rangle_p}_{\text{drag production (DP)}} + \underbrace{\frac{\varphi}{\rho_p}\left\langle u_f'''\frac{\partial p_f'}{\partial x}\right\rangle_p}_{\text{pressure exchange (PE)}} - \underbrace{\frac{\varphi}{\rho_p}\left\langle u_f'''\frac{\partial \sigma_{f,1i}'}{\partial x_i}\right\rangle_p}_{\text{viscous exchange (VE)}}, \tag{6.5}$$

Where u_p is the particle-phase velocity in an Eulerian frame of reference. Here, $\langle(\cdot)\rangle_p = \langle\varepsilon_p(\cdot)\rangle/\langle\varepsilon_p\rangle$. Fluctuations about PA terms are denoted with a double prime. In a similar fashion, the PA operator in the fluid phase is defined as $\langle(\cdot)\rangle_f = \langle\varepsilon_f(\cdot)\rangle/\langle\varepsilon_f\rangle$. Fluctuations about the PA fluid velocity are given by $\boldsymbol{u}_f''' = \boldsymbol{u}_f(\boldsymbol{x},t) - \langle\boldsymbol{u}_f\rangle_f$. With this, the fluid-phase TKE is given by $k_f = \langle\boldsymbol{u}_f'''\cdot\boldsymbol{u}_f'''\rangle_f/2$.

It is notable that all the terms appearing on the right hand side of (6.5) are unclosed and require modeling. This work has already been carried out using sparse regression exclusively in Chapter IV [59]. Here, we select the drag production term to demonstrate the present methodology. This term is chosen due to its importance in this class of flows. In the absence of mean shear, it is the only source of fluid-phase TKE. Additionally, all components of the drag production tensor are identically zero, except for the contribution in the gravity aligned direction. This condition often presents challenges for modeling.

6.4 Results and discussion

We now demonstrate the modeling methodology presented in Section 6.2 on the multiphase case study summarized in Section 6.3, focusing on the drag production term, $\mathcal{R}^{\mathrm{DP}}$, in particular. Here, we follow the three modeling steps as outlined previously.

In the first step, we conduct modeling of drag production using sparse regression

with embedded invariance and the assumption of constant coefficients to inform the bases that comprise the reduced set, $T^{\subset}$. The model consisting of the reduced basis tensors is given as

$$\mathcal{R}^{\mathrm{DP}} = \beta_1 \mathbb{I} + \beta_2 \hat{\mathbb{U}}_r, \tag{6.6}$$

where $\hat{\mathbb{U}}_r$ is a tensor formulated using the mean slip velocity between the phase, $\mathbb{I}$ is the identity tensor and the coefficients, β_1 and β_2, have functional dependency upon configuration parameters that are unknown and cannot be informed by physics-based reasoning.

Next, we evaluate the ideal *constant* coefficients for each unique configuration studied, by conducting OLS and allowing the coefficients, β_1 and β_2 to take on unique values for each configuration. In other words, the values of β_1 and β_2 associated with the case for $\langle \varepsilon_p \rangle = 0.001$ and $g = 0.8$ need not be the same as the values for $\langle \varepsilon_p \rangle = 0.05$ and $g = 2.4$.

As described in Section 6.2, the ideal coefficients are arranged into two vectors: one for each of the basis tensors, $\mathbb{I}$ and $\hat{\mathbb{U}}_r$. Each vector of ideal coefficients is used as input, along with the associated parameters and invariants, to the GEP algorithm [199]. Here, the GEP algorithm selects models that reduce the R^2 between the ideal coefficient values and the candidate models, which are all functions of the parameters and invariants. This effectively collapses the vector of ideal coefficients to a single, compact algebraic expression.

The resultant model learned from this methodology is given as

$$\begin{aligned} \mathcal{R}^{\mathrm{DP}} = &\left(0.258\varphi + (0.03\varphi)^3 + 1.9\frac{\langle \varepsilon_p \rangle}{\mathcal{S}^{(2)}}\right)\mathbb{I} + \\ &\left(1.9\varphi - 5.8\varphi^{1/2}\right)\hat{\mathbb{U}}_r \end{aligned} \tag{6.7}$$

where $\mathcal{S}^{(2)}$ is a principal invariant, defined as $\mathrm{tr}\left(\hat{\mathbb{U}}_r \hat{\mathbb{R}}_f \hat{\mathbb{R}}_p^2\right)$, and the basis tensor, $\hat{\mathbb{U}}_r$ is defined by the normalized slip tensor. This slip tensor is given as the outer product

of the mean slip velocity, $(\langle \boldsymbol{u}_p \rangle_p - \langle \boldsymbol{u}_f \rangle_f) \otimes (\langle \boldsymbol{u}_p \rangle_p - \langle \boldsymbol{u}_f \rangle_f)$. The two other basis tensors, $\hat{\mathbb{R}}_f$ and $\hat{\mathbb{R}}_p$, are the anisotropic stress tensors associated with the fluid and particle phase, respectively. In terms of solution variables, the mean phase velocities are solved by associated momentum equations and the Reynolds stresses are informed by transport equations in the multiphase RANS equations [see 59]. This model has an error of 0.012, where the error is defined as

$$\epsilon = \frac{||\mathbb{D} - \mathbb{T}\hat{\beta}||_2^2}{||\mathbb{D}||_2^2}. \tag{6.8}$$

This is comparable performance to the model learned using sparse regression exclusively ($\epsilon = 0.013$), however, the proposed method does not require a manual selection of trial functions for the coefficients, thus making it far more efficient from a modeling perspective.

As a counter argument to the blended modeling approach, we also allowed the GEP algorithm to learn the full model (i.e., the mean values of drag production, all 24 basis tensors, the principal invariants and configuration parameters from the Euler–Lagrange simulations were provided to the GEP algorithm as input). The learned model is given as

$$\begin{aligned} \mathcal{R}^{\mathrm{DP}} =& 24.4\hat{\mathbb{U}}_r + 30.4e^{-\hat{\mathbb{R}}_p} - 1.42\left(\hat{\mathbb{U}}_r\hat{\mathbb{R}}_p + (\hat{\mathbb{U}}_r\hat{\mathbb{R}}_p)^{\mathrm{T}}\right)^{1/2} + \\ & 1.41\times 10^5\left(\hat{\mathbb{U}}_r^2\hat{\mathbb{R}}_f + (\hat{\mathbb{U}}_r^2\hat{\mathbb{R}}_f)^{\mathrm{T}}\right)^2\langle\varepsilon_p\rangle^2 - 30.4, \end{aligned} \tag{6.9}$$

with associated error 0.13 (an order of magnitude higher than the blended or sparse regression only approach). This degradation in performance can be attributed to the fact that the model now depends upon $\hat{\mathbb{R}}_p$ and $\hat{\mathbb{R}}_f$, the particle and fluid-phase Reynolds stress tensors, in addition to $\hat{\mathbb{U}}_r$. On a fundamental level, since the drag production is a gravity-based phenomenon (i.e., TKE is generated *solely* due to the presence of gravity in this configuration), we can anticipate that $\hat{\mathbb{U}}_r$ would be the

predominant tensor from the basis for describing the physics. Additionally, since the Reynolds stresses contain nonzero diagonal entries, including these terms makes it difficult to drive the cross stream directions of the drag production model to zero. Finally, and perhaps most importantly, GEP does not enforce the relation that resultant model be linear with respect to the basis tensors. The stipulation of linearity in the basis tensors is critical for ensuring form invariance in the resultant model and for ensuring a physics-based description of the data, as described by (6.1). These results suggest that sparse regression and GEP are both needed in order to select a minimal set of tensors to describe physics and automate the complex dependencies of the coefficients when physical intuition cannot guide this process.

CHAPTER VII

Conclusions

7.1 Summary of achievements

This work summarizes several key achievements: (1) The quantification of the role heterogeneity plays on heat transfer and thermochemical conversion, (2) The development of a modeling methodology capable of producing multiphase turbulence closures that are accurate across regimes and (3) Several proposed models, for both single phase as well as two–way coupled, gas-solid turbulence.

In Chapter II, an Eulerian-Lagrangian study was conducted to examine the effect of heterogeneity on the thermochemical conversion of biomass to biofuel. This work found that multiphase behavior (i.e. the formation of clusters) results in a delay of heat transfer and conversion of biomass to biofuel. Stemming from this, at typical residence times, it was found that a model assuming a uniform particle phase and perfect mixing resulted in a 33% over prediction of biofuel yield.

In Chapter III, a modeling framework was developed by extending and modifying the sparse regression strategy employed in Brunton et al. [38] to turbulence modeling for the first time. This method has the following key features:

1. Consistency with the concept that any tensor can be represented as a linear function of basis tensors and coefficients, which also ensures the resultant model is form invariant.

2. Results in a *compact*, algebraic model, due to the sparsity promoting nature of the cost function, that can be formulated 'from scratch'–without reliance on another model as a starting point.

3. Easily interpretable and transportable to other users and forward solvers, owing to the compact, algebraic nature of the models.

Additionally, the sparse regression modeling framework was validated for synthetic data, demonstrating that it can accurately return a *known* model, even when artificial noise was applied. It was then employed to learn new models for canonical turbulent flows: Homogeneous free shear turbulence and turbulence over periodic hills. In each of these configurations, sparse regression learned a model that was more accurate than the existing LEVM model and had comparable accuracy with NN-based models. However, the models learned using sparse regression maintained interpretability, invariance and ease of use and dissemination. Further, this work trained models both for algebraic closure of the anisotropic stress tensor as well as models for an unclosed term appearing in the transport equation of the Reynolds stresses.

Finally, this work demonstrated that the sparse regression methodology was robust to both noise and sparsity in the training dataset, suggesting that it would be successful at informing closures from experimental data, and was accurate at predicting flows outside of it's training set (e.g., flow over a backward facing step).

In Chapter IV, the sparse regression modeling frame work from Chapter III was employed to develop models for the terms appearing in the multiphase transport equations for the Reynolds stresses. In this work, a minimal invariant tensor basis was developed to describe the fluid- and particle-phase Reynolds stresses for the first time. Additionally, models for each of the unclosed terms appearing in the transport equations was proposed, along with closures for the transport of turbulent kinetic energy and dissipation in both phases. These models demonstrated excellent performance when tested on a temporally evolving set of test datasets, especially

considering no other models had previously existed that displayed this degree of accuracy across the considered range of mass loadings.

In Chapter V, an Eularian–Lagrangian framework was employed to investigate the effect of clustering on the thermal entrance length. This work formally defined the thermal entrance length for gas-solid flows for the first time. Further, the complex dependence of thermal entrance length on volume fraction and Péclet; number was quantified and scaling laws for both correlated (clustered) and uncorrelated (uniform) particle phases were proposed for the first time. The one-dimensional, two-fluid heat equation was derived, and the relative importance of each term was assessed. This gave rise to the identification of the fluid drift temperature as the sole unclosed term of importance for correlated flows. Given this result, scaling arguments and gene expression programming were leveraged to propose a new closure capable of predicting thermal behavior in coupled, gas-solid flows.

In Chapter VI, a modified version of the the sparse regression methodology presented in Chapter III was proposed. This new methodology blends sparse regression with gene expression programming and is the first blended modeling strategy of its kind. By leveraging these two techniques, the important properties of sparse regression, namely the integration of knowledge of physics and model properties such as frame invariance, are maintained while the formulation of the functional form of the coefficients is automated by the GEP algorithm. Since the functional dependence of the coefficients can rarely be informed by physical inferences, offloading this work to a naive, heuristic algorithm makes the modeling process more streamlined and efficient.

The contributions listed above are also in print or have been submitted for publication in the following:

1. **Beetham, S.**, Capecelatro, J. (2019) Biomass pyrolysis in fully-developed turbulent riser flow. *Renewable Energy.* 140, 751-760.

2. **Beetham, S.**, Capecelatro, J., (2020) Formulating turbulence closures using

sparse regression with embedded form invariance. *Physical Review Fluids*. 5, 084611.

3. **Beetham, S.**, Fox, R.O., Capecelatro, J., (2021) Sparse identification of multiphase turbulence closures for coupled fluid-particle flows. *Journal of Fluid Mechanics*. 914, A11.

4. **Beetham, S.**, Lattanzi, A., Capecelatro, J., (2021) On the thermal entrance length of moderately dense gas-particle flows. *International Journal of Heat and Mass Trasnfer*. sumbitted.

5. **Beetham, S.**, Capecelatro, J., (2022) Multiphase turbulence modeling using sparse regression and gene expression programming. *Proceedings of the International Union on Theoretical and Applied Mechanics Symposium*. submitted

Also published during my PhD studies, but not contained within this thesis is the following manuscript:

6. **Verner, S. N.**, Garikipati, K. (2018) A computational study of the mechanisms growth-driven folding patterns on shells, with application to the developing brain. *Extreme Mechanics Letters* 18, 58-69.

7.1.1 A note on the universality of learned models

A principal challenge of any data-driven modeling effort, is assessing the applicability, or universality, of the learned model. While it cannot be known *a priori* if any model will be accurate across *every* flow condition, a few steps can be taken to assess the goodness of a model or predict reach of its predictability. First, because the learned models here are in compact, algebraic form, model properties, such as stability or equilibrium values, can be assessed. These model attributes can identify conditions under which a model may require further training/refinement or regimes

in which it is likely to perform well. Further, integration into existing solvers can be treated in such a way that the solver uses the learned model in regimes that in which it has been trained and tested. In all other regimes, a baseline model can be used. This strategy mitigates the risk of incorporating new models in regions far outside its scope of training.

7.2 Looking forward to future work

7.2.1 Developing multiphase models from a wider parameter set

The focus of the work in Chapters IV and VI was on gas-particle flows, with a fixed density ratio of 1000. In order to formulate models with a wider range of applicability, future work will expand the density ratio parameter to include liquid-gas flows, which have a much smaller density ratio. In this work, the blended sparse regression and GEP methodology will be leveraged to extend the multiphase RANS models proposed in Chapters IV and VI to multiphase flows ranging from bubbly flow to gas-solid flow.

7.2.2 Penalizing sparse regression to ensure model stability

In this work, the sparse regression modeling methodology was demonstrated to be able to embed important model attributes, such as form invariance. To extend this notion, another important model quality is stability and the ability to embed this property during the modeling process will be instrumental in ensuring learned models are well posed in the context of a forward solver. In order to achieve this, an additional penalty will be added to the sparse regression cost function which minimizes the spectral radius of the resultant model, i.e.,

$$\hat{\beta} = \min_{\beta} ||\mathbb{D} - \mathbb{T}\beta||_2^2 + \lambda_1||\beta||_1 + \lambda_2||\rho(\mathbb{A})||_2^2, \tag{7.1}$$

where $\rho(\mathbb{A})$ denotes the spectral radius of $\mathbb{A}$, the operator matrix associated with the linearized system of equations for the forward solver (e.g., $\mathbb{A}\boldsymbol{x} = \boldsymbol{b}$). Here, the formulation of $\mathbb{A}$ is dependent upon discretization choices in the forward solver and assumes that the model $\mathbb{T}\beta$ is accounted for as a part of the system and not as a forcing term appearing in $\boldsymbol{b}$. In this cost function, two optimization parameters appear (λ_1 and λ_2) and can be tuned independently.

7.2.3 Formulating closures for near-wall effects and bi-disperse flows

In order to increase the utility of the closures presented in this work, the modeling methodologies presented here will be leveraged to formulate models that can accommodate more realistic flow conditions, as seen in a configuration such as a circulating fluidized bed reactor (see Chapter II) such as near wall effects and a bi-disperse particle phase. The approach will follow what was presented in this work: highly resolved Eulerian-Lagrangian simulations will be leveraged to isolate key physics (e.g., particle laden channel and duct flow can be utilized to isolate near wall effects and homogeneous CIT with bi-disperse particles can isolate the physics associated with inhomogeneity in particle parameters). Then, this data can be used as input to the modeling methodologies set forth in this work.

APPENDICES

APPENDIX A

Non-dimensionalization of the Heat Equation

The volume-filtered fluid temperature equation is given by

$$\frac{\partial}{\partial t}\left(\varepsilon_f \rho_f C_{p,f} T_f\right) + \nabla \cdot \left(\varepsilon_f \rho_f C_{p,f} \boldsymbol{u}_f T_f\right) = \varepsilon_f \nabla \cdot \left(\kappa_f \nabla T_f\right) + \mathcal{Q}_{\text{inter}}, \tag{A.1}$$

where,

$$\mathcal{Q}_{\text{inter}} = -\sum_{i=1}^{N_p} \mathcal{G}\left(|\boldsymbol{x} - \boldsymbol{x}_p^{(i)}|\right) q_{\text{heat}}^{(i)}, \tag{A.2}$$

and

$$q_{\text{heat}}^{(i)} = V_p \left[\frac{6\kappa_f \text{Nu}}{d_p^2}\left(T_f[x_p^{(i)}] - T_p^{(i)}\right)\right]. \tag{A.3}$$

This results in the following expression,

$$\begin{aligned}\frac{\partial}{\partial t}\left(\varepsilon_f \rho_f C_{p,f} T_f\right) + &\nabla \cdot \left(\varepsilon_f \rho_f C_{p,f} \boldsymbol{u}_f T_f\right) = \\ &\varepsilon_f \nabla \cdot \left(\kappa_f \nabla T_f\right) + \frac{6\varepsilon_p \kappa_f \text{Nu}}{d_p^2}\left(T_f - T_p\right).\end{aligned} \tag{A.4}$$

We first non-dimensionalize temperature using $\theta = (T - T_{p,0})/(T_{f,0} - T_{p,0})$. This

yields,

$$\frac{\partial}{\partial t}\left(\varepsilon_f \rho_f C_{p,f} \theta_f\right) + \nabla \cdot \left(\varepsilon_f \rho_f C_{p,f} \boldsymbol{u}_f \theta_f\right) = \varepsilon_f \nabla \cdot \left(\kappa_f \nabla \theta_f\right) + \frac{6 \varepsilon_p \kappa_f \mathrm{Nu}}{d_p^2}\left(\theta_f - \theta_p\right). \tag{A.5}$$

Now, we divide through by $C_{p,f}\rho_f$ and make the change of variable $\hat{\boldsymbol{x}} = \boldsymbol{x}/d_p$, which gives

$$\frac{\partial}{\partial t}\left(\varepsilon_f \theta_f\right) + \left(\frac{1}{d_p}\right)\frac{\partial}{\partial \hat{\boldsymbol{x}}}\left(\varepsilon_f \boldsymbol{u}_f \theta_f\right) = \frac{\varepsilon_f}{\rho_f C_{p,f}}\left(\frac{1}{d_p}\right)\frac{\partial}{\partial \hat{\boldsymbol{x}}}\left(\kappa_f \left(\frac{1}{d_p}\right)\frac{\partial}{\partial \hat{\boldsymbol{x}}}\theta_f\right) + \frac{6 \varepsilon_p \kappa_f \mathrm{Nu}}{\rho_f C_{p,f} d_p^2}\left(\theta_f - \theta_p\right). \tag{A.6}$$

Multiplying by (d_p/u_{bulk}) gives rise to

$$\frac{d_p}{u_{\mathrm{bulk}}}\frac{\partial}{\partial t}\left(\varepsilon_f \theta_f\right) + \frac{d_p}{u_{\mathrm{bulk}}}\left(\frac{1}{d_p}\right)\frac{\partial}{\partial \hat{\boldsymbol{x}}}\left(\varepsilon_f \boldsymbol{u}_f \theta_f\right) = \frac{d_p}{u_{\mathrm{bulk}}}\varepsilon_f\left(\frac{1}{d_p}\right)\frac{\partial}{\partial \hat{\boldsymbol{x}}}\left(\alpha_f \left(\frac{1}{d_p}\right)\frac{\partial}{\partial \hat{\boldsymbol{x}}}\theta_f\right) + \frac{6 \varepsilon_p \alpha_f \mathrm{Nu}}{d_p u_{\mathrm{bulk}}}\left(\theta_f - \theta_p\right), \tag{A.7}$$

where we notice that a Péclet number arises in both terms on the right hand side as,

$$\frac{d_p}{u_{\mathrm{bulk}}}\frac{\partial}{\partial t}\left(\varepsilon_f \theta_f\right) + \frac{\partial}{\partial \hat{\boldsymbol{x}}}\left(\frac{\varepsilon_f \boldsymbol{u}_f \theta_f}{u_{\mathrm{bulk}}}\right) = \varepsilon_f \frac{\partial}{\partial \hat{\boldsymbol{x}}}\left(\frac{1}{\mathrm{Pe}}\frac{\partial}{\partial \hat{\boldsymbol{x}}}\theta_f\right) + \frac{6\mathrm{Nu}}{\mathrm{Pe}}\varepsilon_p\left(\theta_f - \theta_p\right). \tag{A.8}$$

Finally, we define a timescale using d_p/u_{bulk} and a non-dimensional time as $\hat{t} = t/(d_p/u_{bulk})$. Making this change of variable yields

$$\frac{\partial}{\partial \hat{t}}\left(\varepsilon_f \theta_f\right) + \frac{\partial}{\partial \hat{\boldsymbol{x}}}\left(\frac{\varepsilon_f \boldsymbol{u}_f \theta_f}{u_{bulk}}\right) = \varepsilon_f \frac{\partial}{\partial \hat{\boldsymbol{x}}}\left(\frac{1}{Pe}\frac{\partial}{\partial \hat{\boldsymbol{x}}}\theta_f\right) + \frac{6\mathrm{Nu}}{\mathrm{Pe}}\varepsilon_p\left(\theta_f - \theta_p\right). \tag{A.9}$$

Finally, a non-dimensional velocity is defined as $\hat{\boldsymbol{u}}_f = \boldsymbol{u}_f / u_{bulk}$ and the final non-dimensional equation is given by

$$\frac{\partial}{\partial \hat{t}}\left(\varepsilon_f \theta_f\right) + \frac{\partial}{\partial \hat{\boldsymbol{x}}}\left(\varepsilon_f \hat{\boldsymbol{u}}_f \theta_f\right) = \varepsilon_f \frac{\partial}{\partial \hat{\boldsymbol{x}}}\left(\frac{1}{Pe}\frac{\partial}{\partial \hat{\boldsymbol{x}}}\theta_f\right) + \frac{6\mathrm{Nu}}{\mathrm{Pe}}\varepsilon_p\left(\theta_f - \theta_p\right), \tag{A.10}$$

and can be reorganized as

$$\mathrm{Pe}\frac{\partial}{\partial \hat{t}}\left(\varepsilon_f \theta_f\right) + \mathrm{Pe}\frac{\partial}{\partial \hat{\boldsymbol{x}}}\left(\varepsilon_f \hat{\boldsymbol{u}}_f \theta_f\right) = \varepsilon_f \frac{\partial^2 \theta_f}{\partial \hat{\boldsymbol{x}}^2} + 6\mathrm{Nu}\varepsilon_p\left(\theta_f - \theta_p\right). \tag{A.11}$$

The particle phase heat equation (Eq. (2.25)) can be similarly nondimensionalized. First, the particle phase heat equation is rewritten in the Eulerian sense by conducting a change of frame from the Lagrangian particle heat equation and projecting it to the Eulerian grid. This Eulerian representation is given as,

$$\rho_p C_{p,p}\left(\frac{\partial\left(\varepsilon_p T_p\right)}{\partial t} + \nabla\cdot\left(\varepsilon_p \boldsymbol{u}_p T_p\right)\right) = \varepsilon_p \kappa_f \nabla^2 T_f + \frac{6\varepsilon_p \kappa_f \mathrm{Nu}}{d_p^2}\left(T_f - T_p\right). \tag{A.12}$$

In the same manner as the fluid heat equation, the particle heat equation is nondimensionalized yielding

$$\chi\ \mathrm{Pe}\frac{\rho_p}{\rho_f}\left[\frac{\partial\left(\varepsilon_p \theta_p\right)}{\partial \hat{t}} + \frac{\partial\left(\varepsilon_p \hat{\boldsymbol{u}}_p \theta_p\right)}{\partial \hat{\boldsymbol{x}}}\right] = \varepsilon_p \frac{\partial^2 \theta_f}{\partial \hat{\boldsymbol{x}}^2} + 6\mathrm{Nu}\varepsilon_p\left(\theta_f - \theta_p\right) \tag{A.13}$$

where, χ is the ratio of heat capacities.

APPENDIX B

Development of the 1D Heat Equation

The fluid-phase energy equation is given as

$$\frac{\partial}{\partial t}\left(\varepsilon_f \rho_f C_{p,f} T_f\right) + \nabla \cdot \left(\varepsilon_f \rho_f C_{p,f} \boldsymbol{u}_f T_f\right) = \varepsilon_f \nabla \cdot \left(\kappa_f \nabla T_f\right) + \mathcal{Q}_{\text{inter}}. \tag{B.1}$$

We conduct Reynolds averaging with respect to the cross-stream directions and time, and assume negligible effects from thermal diffusion. This results in a much simpler expression, given by

$$C_{p,f}\rho_f \frac{\mathrm{d}}{\mathrm{d}x}\langle \varepsilon_f \boldsymbol{u}_f T_f \rangle = \langle \mathcal{Q}_{\text{inter}} \rangle. \tag{B.2}$$

Note that the interphase heat exchange is given as

$$\mathcal{Q}_{\text{inter}} = -\sum_{i=1}^{N_p} \mathcal{G}\left(|\boldsymbol{x} - \boldsymbol{x}_p^{(i)}|\right) q_{\text{heat}}^{(i)}, \tag{B.3}$$

where,

$$q_{\text{heat}}^{(i)} = V_p \left[\frac{6\kappa_f \text{Nu}}{d_p^2}\left(T_f[x_p^{(i)}] - T_p^{(i)}\right)\right]. \tag{B.4}$$

Substituting these definitions into the heat equations and noting that a phase average

arises due to the volume fraction in the Reynolds average on the left hand side yields

$$C_{p,f}\rho_f\langle\varepsilon_f\rangle\frac{\mathrm{d}}{\mathrm{d}x}\langle\boldsymbol{u}_f T_f\rangle_f = -\left\langle\frac{6\varepsilon_p\kappa_f\mathrm{Nu}}{d_p^2}\left(T_f-T_p\right)\right\rangle \tag{B.5}$$

$$C_{p,f}\rho_f\langle\varepsilon_f\rangle\frac{\mathrm{d}}{\mathrm{d}x}\langle\boldsymbol{u}_f T_f\rangle_f = -\frac{6\kappa_f}{d_p^2}\langle\varepsilon_p\mathrm{Nu}\left(T_f-T_p\right)\rangle \tag{B.6}$$

Both sides can be simplified further. Working first with the left hand side, we expand $\boldsymbol{u}_f$ and T_f using the Phase averaged decompositions, as:

$$C_{p,f}\rho_f\langle\varepsilon_f\rangle\frac{\mathrm{d}}{\mathrm{d}x}\langle(\langle\boldsymbol{u}_f\rangle_f+\boldsymbol{u}_f''')(\langle T_f\rangle_f+T_f''')\rangle_f = -\frac{6\kappa_f}{d_p^2}\langle\varepsilon_p\mathrm{Nu}\left(T_f-T_p\right)\rangle \tag{B.7}$$

$$C_{p,f}\rho_f\langle\varepsilon_f\rangle\frac{\mathrm{d}}{\mathrm{d}x}\left(\langle\langle\boldsymbol{u}_f\rangle_f\langle T_f\rangle_f\rangle_f+\langle\boldsymbol{u}_f'''T_f'''\rangle_f\right) = -\frac{6\kappa_f}{d_p^2}\langle\varepsilon_p\mathrm{Nu}\left(T_f-T_p\right)\rangle \tag{B.8}$$

$$C_{p,f}\rho_f\langle\varepsilon_f\rangle\frac{\mathrm{d}}{\mathrm{d}x}\left(\langle\boldsymbol{u}_f\rangle_f\langle T_f\rangle_f+\langle\boldsymbol{u}_f'''T_f'''\rangle_f\right) = -\frac{6\kappa_f}{d_p^2}\langle\varepsilon_p\mathrm{Nu}\left(T_f-T_p\right)\rangle \tag{B.9}$$

$$C_{p,f}\rho_f\langle\varepsilon_f\rangle\left(\langle\boldsymbol{u}_f\rangle_f\frac{\mathrm{d}\langle T_f\rangle_f}{\mathrm{d}x}+\frac{\mathrm{d}}{\mathrm{d}x}\langle\boldsymbol{u}_f'''T_f'''\rangle_f\right) = -\frac{6\kappa_f}{d_p^2}\langle\varepsilon_p\mathrm{Nu}\left(T_f-T_p\right)\rangle \tag{B.10}$$

Turning now to the right hand side, we notice that a phase-average with respect to the particle phase is present.

$$\begin{aligned}
C_{p,f}\rho_f\langle\varepsilon_f\rangle&\left(\langle\boldsymbol{u}_f\rangle_f\frac{\mathrm{d}\langle T_f\rangle_f}{\mathrm{d}x}+\frac{\mathrm{d}}{\mathrm{d}x}\langle\boldsymbol{u}''T_f''\rangle_f\right)\\
&= -\frac{6\kappa_f}{d_p^2}\left\langle\varepsilon_p\mathrm{Nu}(T_f-T_p)\right\rangle\\
&= -\frac{6\kappa_f\langle\varepsilon_p\rangle}{d_p^2}\left\langle\mathrm{Nu}(T_f-T_p)\right\rangle_p\\
&= -\frac{6\kappa_f\langle\varepsilon_p\rangle}{d_p^2}\left\langle(\langle\mathrm{Nu}\rangle_p+Nu'')(\langle T_f\rangle_p+T_f''-\langle T_p\rangle_p-T_p'')\right\rangle_p\\
&= -\frac{6\kappa_f\langle\varepsilon_p\rangle}{d_p^2}\left[\langle\mathrm{Nu}\rangle_p\left(\langle T_f\rangle_p-\langle T_p\rangle_p\right)+\left(\langle\mathrm{Nu}''T_f''\rangle-\langle\mathrm{Nu}''T_p''\rangle\right)\right]
\end{aligned} \tag{B.11}$$

APPENDIX C

Reynolds-Averaged Contributions to Phase-Averaged Terms

While the phase-averaged equations have mathematical utility as this significantly reduces the number of terms as compared with Reynolds averaging, proposing models in for the phase-averaged equations in the context of the present study requires additional closure for boundary conditions, since the correlation between temperature and volume fraction fluctuations at the inlet to the thermal domain cannot be known *a priori*. To maintain consistency in boundary conditions in comparing models for correlated and uncorrelated flows, we shift to the Reynolds averaged descriptions of the surviving terms in the phase averaged equations. For the fluid-phase, this exercise results in the expression,

$$\underbrace{\langle\hat{u}_f\rangle\frac{\mathrm{d}\langle\theta_f\rangle}{\mathrm{d}\hat{x}}+\frac{\langle\varepsilon_f'\hat{u}_f'\rangle}{\langle\varepsilon_f\rangle}\frac{\mathrm{d}\langle\theta_f\rangle}{\mathrm{d}\hat{x}}+\frac{\langle\hat{u}_f\rangle}{\langle\varepsilon_f\rangle}\frac{\mathrm{d}\langle\varepsilon_f'\theta_f'\rangle}{\mathrm{d}\hat{x}}+\frac{\langle\varepsilon_f'\hat{u}_f'\rangle}{\langle\varepsilon_f\rangle^2}\frac{\mathrm{d}\langle\varepsilon_f'\theta_f'\rangle}{\mathrm{d}\hat{x}}}_{\text{Convection}}= \tag{C.1}$$

$$-\frac{6\langle\varepsilon_p\rangle\widetilde{Nu}}{\mathrm{Pe}\langle\varepsilon_f\rangle}\left[\underbrace{\langle\theta_f\rangle-\langle\theta_p\rangle+\frac{\langle\varepsilon_f'\theta_f'\rangle}{\langle\varepsilon_f\rangle}-\frac{\langle\varepsilon_p'\theta_p'\rangle}{\langle\varepsilon_p\rangle}}_{\text{Term 2}}+\underbrace{\frac{\langle\varepsilon_p'\theta_f'\rangle}{\langle\varepsilon_p\rangle}-\frac{\langle\varepsilon_f'\theta_f'\rangle}{\langle\varepsilon_f\rangle}}_{\text{Term3}}\right]$$

where $\widetilde{Nu}$ denotes the Nusslet number computed using the correlation proposed by

Sun and Zhu [171] and mean quantities as arguments. As shown in the detailed panels of Fig. C.1, all of the unclosed Reynolds averaged terms are null, except for the cross correlation between particle volume fraction and the fluid-phase temperature fluctuations arising from Term 3, as would be expected from Guo and Capecelatro [56]. Thus, the simplified Reynolds averaged equation is given as

$$\langle \hat{u}_f \rangle \frac{\mathrm{d}\langle \theta_f \rangle}{\mathrm{d}\hat{x}} = -\frac{6\langle \varepsilon_p \rangle \widetilde{Nu}}{\mathrm{Pe}\langle \varepsilon_f \rangle} \left[\langle \theta_f \rangle - \langle \theta_p \rangle + \frac{\langle \varepsilon_p' \theta_f' \rangle}{\langle \varepsilon_p \rangle} \right]. \tag{C.2}$$

This result points to the fact that cross-correlations between volume fraction and temperature shift the phase averaged temperature from the Reynolds averaged temperature (e.g., $\langle \theta_f(\hat{x}) \rangle_f = \langle \theta_f(\hat{x}) \rangle + \langle \varepsilon_f' \theta_f' \rangle$), where in these configurations the cross correlations are constant with respect to $\hat{x}$. Thus, the model proposed herein is suitable for use in simulations for which the solution variables are phase-averaged *or* Reynolds averaged (see Fig. 5.8). This also implies that the proposed model is appropriate for use in a general two-fluid solver in which the hydrodynamics and thermodynamics evolve simultaneously. Of course, in this situation, additional closures are required for the fluid and particle momentum equations in order to capture cross correlations.

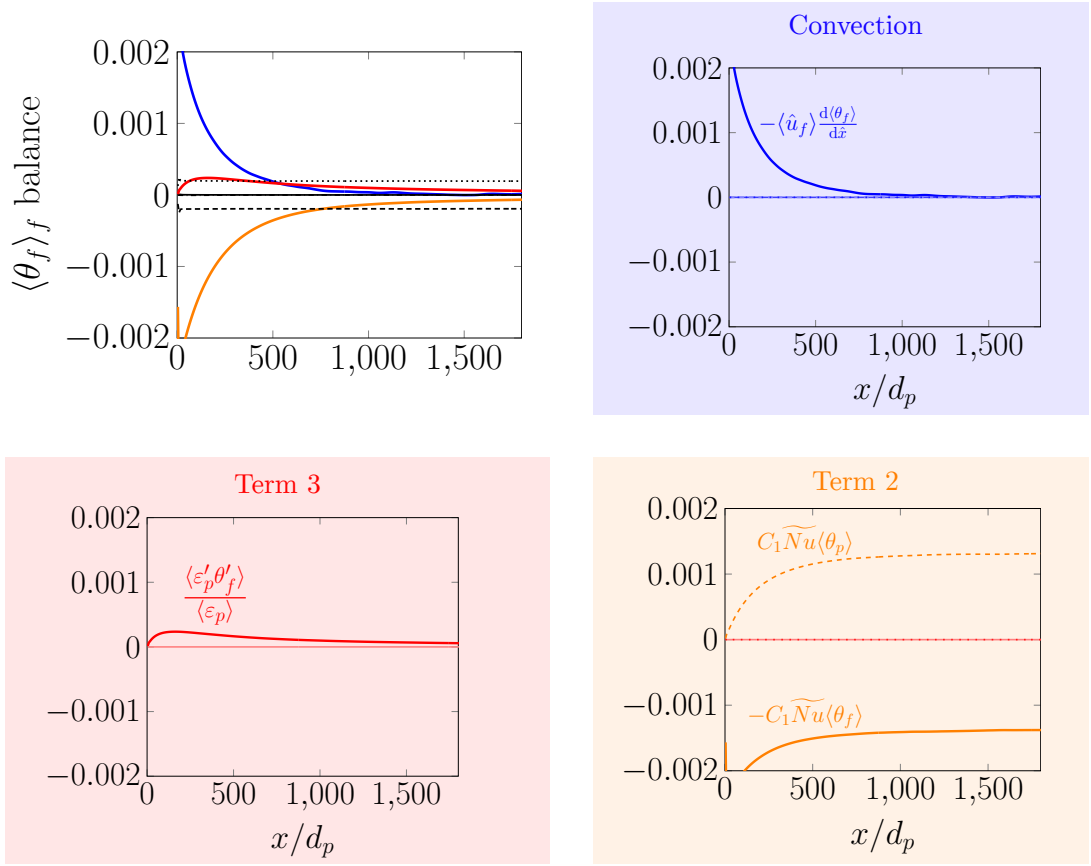


Figure C.1: Upper left: Of the terms appearing in the phase-averaged, fluid-phase energy equation, three dominate the thermal behavior: Convection (blue), Term 2 (orange) and Term 3 (red). Terms 4 (black, dotted) and 5 (black, dashed) are nonzero, but balance each other exactly. Colored breakout panels of the three dominate terms detail the contributions to each of these in terms of *Reynolds averaged* quantities. Of the nonzero terms, only $\langle\varepsilon'_p\theta'_f\rangle$ requires modeling.

BIBLIOGRAPHY

BIBLIOGRAPHY

[1] J. Towns, T. Cockerill, M. Dahan, I. Foster, K. Gaither, A. Grimshaw, V. Hazlewood, S. Lathrop, D. Lifka, G. D. Peterson, R. Roskies, J. R. Scott, and N. Wilkins-Diehr. Xsede: Accelerating scientific discovery. *Computing in Science & Engineering*, 16:62–74, 2014.

[2] A. Kolmogorov. Dissipation of energy in locally isotropic turbulence. *Doklady Akademii Nauk SSSR*, 32:16–18, 1941.

[3] Said Elgobashi. An updated classification map of particle-laden turbulent flows. In *IUTAM Symposium on Computational Approaches to Multiphase Flow*, pages 3–10. Springer, 2006.

[4] F. Shaffer, B. Gopalan, R. W. Breault, R. Cocco, S. B. Karri, R. Hays, and T. Knowlton. High speed imaging of particle flow fields in CFB risers. *Powder Technology*, 242:86–99, 2013.

[5] S. Tenneti and S. Subramaniam. Particle-resolved direct numerical simulation for gas-solid flow model development. *Annu. Rev. Fluid Mech.*, 46:199–230, 2014.

[6] S. Beetham and J. Capecelatro. Biomass pyrolysis in fully-developed turbulent riser flow. *Renewable Energy*, 140:751–760, 2019.

[7] C. Di Blasi. Numerical simulation of cellulose pyrolysis. *Biomass and Bioenergy*, 7:87–98, 1994.

[8] C.A. Koufopanosi, N. Papayannakos, G. Maschio, and A. Lucchesi. Modelling of the pyrolysis of biomass particles. studies on kinetics, thermal and heat transfer effects. *The Canadian Journal of Chemical Engineering*, 69:907–915, 1991.

[9] L.J. Curtis and D.J Miller. Transport model with radiative heat transfer for rapid cellulose pyrolysis. *Industrial & Engineering Chemistry Research*, 27: 1775–1783, 1988.

[10] M. Breuer, N. Peller, C. Rapp, and M. Manhart. Flow over periodic hills–numerical and experimental study in a wide range of Reynolds numbers. *Computers & Fluids*, 38:433–457, 2009.

[11] P. Moin and J. Kim. Tackling turbulence with supercomputers. *Scientific American*, 276(1):62–68, 1997. ISSN 00368733, 19467087. URL `http://www.jstor.org/stable/24993565`.

[12] S. Jovic and D. M. Driver. Backward-facing step measurements at low reynolds number, $\mathrm{Re}_h = 5000$. *NASA Technical Memorandum*, 108807, 1994.

[13] Ch. Rapp and M. Manhart. Flow over periodic hills: an experimental study. *Experimental Fluids*, 51:247–269, 2011.

[14] O. Reynolds. On the dynamical theory of incompressible viscous fluids and the determination of the criterion. *Philosophical Transactions of the Royal Society A*, 186, 1895.

[15] Joseph Boussinesq. *Thōrie analytique de la chaleur mise en harmonie avec la thermodynamique et avec la thōrie mcanique de la lumire: Refroidissement et chauffement par rayonnement, conductibilit des tiges, lames et masses cristallines, courants de convection, thōrie mcanique de la lumire. 1903. xxxii, 625,[1] p*, volume 2. Gauthier-Villars, 1903.

[16] A. J. M. Spencer and R. S. Rivlin. The theory of matrix polynomials and its application to the mechanics of isotropic continua. *Archive for Rational Mechanics and Analysis*, 2:309–336, 1958.

[17] S. B. Pope. Turbulent flows. *Cambridge University Press*, 2000.

[18] B. E. Launder, G. J. Reese, and W. Rodi. Progress in the development of a Reynolds-stress turbulence closure. *Journal of Fluid Mechanics*, 68:537–566, 1975.

[19] L. Prandtl. Bericht uber untersuchungen zur ausgebildeten turbulenz. *Zs. Angew. Math. Mech.*, 2, 1925.

[20] J. Smaogorinsky. General circulation experiments with the primitive eequations. 1. the basic experiment. *Monthly Weather Review*, 91(3):99–164, 1963.

[21] Karthik Duraisamy, Gianluca Iaccarino, and Heng Xiao. Turbulence modeling in the age of data. *Annual Review of Fluid Mechanics*, 51:357–377, 2019.

[22] B. Tracey, K. Duraisamy, and J. J. Alonso. A machine learning strategy to assist turbulence model development. *AIAA Paper*, 1287, 2015.

[23] M. Milano and P. Koumoutsakos. Neural network modeling for near wall turbulent flow. *Journal of Computational Physics*, 182:1–26, 2002.

[24] C. Lu. Artificial neural network for behavior learning from meso-scale simulations, application to multi-scale multimaterial flows. *PhD thesis*, 2010.

[25] E. Rajabi and M. R. Kavianpour. Intelligent prediction of turbulent flow over backward-facing step using direct numerical simulation data. *Engineering Applications of Computational Fluid Mechanics*, 6(4):490–503, 2012.

[26] K. Duraisamy and P. A. Durbin. Transition modeling using data driven approaches. *Proceedings of the Summer Program*, page 427, 2014.

[27] M. Ma, J. Lu, and G. Tryggvason. Using statistical learning to close two-fluid multiphase flow equations for bubbly flows in vertical channels. *International Journal of Multiphase Flow*, 85:336–347, 2016.

[28] J. Ling, A. Kurzawski, and J. Templeton. Reynolds averaged turbulence modeling using deep neural networks with embedded invariance. *Journal of Fluid Mechanics*, 807:155–166, 2016.

[29] M. Bode, M. Gauding, K. Kleinheinz, and H. Pitsch. Deep learning at scale for subgrid modeling in turbulent flows: regression and reconstruction. *arXiv:1910.00928v1*, 2019.

[30] W. Liu and J. Fang. Iterative framework of machine-learning based turbulence modeling for Reynolds-averaged Navier-Stokes simulations. *arXiv:1910.01232v1*, 2019.

[31] M. Schmelzer and R. Dwight. Data-driven deterministic symbolic regression of nonlinear stress-strain relation for RANS turbulence modelling. *AIAA Aviation Forum*, 2018.

[32] M. Schmelzer, R. Dwight, and P. Cinnella. Discovery of algebraic Reynolds-stress models using sparse symbolic regression. *Flow, Turbulence and Combustion*, 104:579–603, 2020.

[33] J. Weatheritt and R. D. Sandberg. Improved junction body flow modeling through data-driven symbolic regression. *Journal of Ship Research*, 63(4):283–293, 2019.

[34] C. Lav, R. D. Sandberg, and J. Philip. A framework to develop data-driven turbulence models for flows with organized unsteadiness. *Journal of Computational Physics*, 383:148–165, 2019.

[35] Zhao Y., Akolekar H.D., Weatheritt J., Vittorio M., and Sandberg R.D. Turbulence model development using cfd-driven machine learning. *submitted to Journal of Fluid Mechanics*, 2019.

[36] J. X. Wang, J.L. Wu, and H. Xiao. A physics informed machine learning approach for reconstructing Reynolds stress modeling discrepancies based on DNS data. *Physical Review Fluids*, 2, 2017.

[37] J. Wu, H. Xiao, and E. Paterson. Physics-informed machine learning for augmenting turbulence models: a comprehensive framework. *submitted to: Physical Review Fluids*, 2018.

[38] S. L. Brunton, J. L. Proctor, and J. N. Kutz. Discovering governing equations from data by sparse identification of nonlinear dynamical systems. *Proceedings of the National Academy of Sciences*, 113(15):3932–3937, 2016.

[39] R. O. Fox. On multiphase turbulence models for collisional fluid–particle flows. *Journal of Fluid Mechanics*, 742:368–424, 2014.

[40] J. Capecelatro, O. Desjardins, and R. O. Fox. Numerical study of collisional particle dynamics in cluster-induced turbulence. *Journal of Fluid Mechanics*, 747:R2 1–13, 2014.

[41] J. Capecelatro, O. Desjardins, and R. O. Fox. On fluid-particle dynamics in fully-developed cluster-induced turbulence. *Journal of Fluid Mechanics*, 780: 578–635, 2015.

[42] J. Capecelatro, O. Desjardins, and R. O. Fox. Strongly–coupled gas–particle flows in vertical channels. Part II: Turbulence modeling. *Physics of Fluids*, 28: 1–22, 2016.

[43] D. Veynante and L. Vervisch. Turbulent combustion modeling. *Progress in Energy and Combustion Science*, 28(3):193–266, 2002.

[44] H. Pitsch. Large-eddy simulation of turbulent combustion. *Annu. Rev. Fluid Mech.*, 38:453–482, 2006.

[45] J. L. Sinclair and R. Jackson. Gas–particle flow in a vertical pipe with particle-particle interactions. *AIChE Journal*, 35(9):1473–1486, 1989.

[46] S. Dasgupta, R. Jackson, and S. Sundaresan. Turbulent gas–particle flow in vertical risers. *AIChE Journal*, 40(2):215–228, 1994.

[47] S. Sundaram and L. R. Collins. Spectrum of density fluctuations in a particle-fluid system-I. Monodisperse spheres. *International Journal of Multiphase Flow*, 20(6):1021–1037, 1994.

[48] J. Cao and G. Ahmadi. Gas–particle two–phase turbulent flow in a vertical duct. *International Journal of Multiphase Flow*, 21(6):1203–1228, 1995.

[49] S. Dasgupta, R. Jackson, and S. Sundaresan. Gas–particle flow in vertical pipes with high mass loading of particles. *Powder Technology*, 96:6–23, 1998.

[50] Y. Cheng, Y. Guo, F. Wei, Y. Jin, and W. Lin. Modeling the hydrodynamics of downer reactors based on kinetic theory. *Chemical Engineering Science*, 54: 2019–2027, 1999.

[51] Y.Y. Jiang and P. Zhang. Numerical investigation of slush nitrogen flow in a horizontal pipe. *Chemical Engineering Science*, 73:169–180, 2012.

[52] A. Rao, J. S. Curtis, B. C. Hancock, and C. Wassgren. Numerical simulation of dilute turbulent gas–particle flow with turbulence modulation. *AIChE Journal*, 58:1381–1396, 2012.

[53] Zh. X. Zeng and L. X. Zhou. A two-scale second–order moment particle turbulence model and simulation of dense gas–particle flows in a riser. *Powder Technology*, 162:27–32, 2006.

[54] J. Capecelatro, O. Desjardins, and R. O. Fox. Strongly–coupled gas–particle flows in vertical channels. part i: Reynolds-averaged two–phase statistics. *Physics of Fluids*, 28:1–22, 2016.

[55] A. Innocenti, R. O. Fox, M.V. Salvetti, and S. Chibbaro. A Lagrangian probability-density-function model for collisional turbulent fluid-particle flows. *Journal of Fluid Mechanics*, 862:449–489, 2019.

[56] L. Guo and J. Capecelatro. The role of clusters on heat transfer in sedimenting gas-solid flows. *International Journal of Heat and Mass Transfer*, 132:1217–1230, 2019.

[57] S. Beetham, A. Lattanzi, and J. Capecelatro. On the thermal entrance length of moderately dense gas-particle flows. *International Journal of Heat and Mass Transfer*, page submitted, 2021.

[58] S. Beetham and J. Capecelatro. Formulating turbulence closures using sparse regression with embedded form invariance. *Physical Review Fluids*, 5:084611, 2020.

[59] S. Beetham, R. O. Fox, and J. Capecelatro. Sparse identification of multiphase turbulence closures for coupled fluid-particle flows. *Journal of Fluid Mechanics*, 914, A11, 2021.

[60] S. Beetham and J. Capecelatro. Toward robust turbulence models for strongly-coupled gas-particle flows. *Proceedings of the International Union on Theoretical and Applied Mechanics Symposium*, page submitted, 2021.

[61] M. M. Write, J. A. Satrio, R. C. Brown, D. E. Daugaard, and D. D. Hsu. Techno-economic analysis of biomass fast pyrolysis to transportation fuels. *NREL Technical Report*, NREL/TP-6A20-46586, 2010.

[62] M. S. Mettler, D. G. Vlachos, and P. J. Dauenhauer. Top ten fundamental challenges of biomass pyrolysis for biofuels. *Energy & Environmental Science*, 5:7797–7809, 2012.

[63] AV Bridgwater. Review of fast pyrolysis of biomass and product upgrading. *Biomass and Bioenergy*, 38:68–94, 2012.

[64] A. E. Farrell, R. J. Plevin, B. T. Turner, A. D Jones, M. O'Hare, and D. M. Kammen. Ethanol can contribute to energy and environmental goals. *Science*, 311:506–508, 2006.

[65] M. P. Dudukovic. Frontiers in reactor engineering. *Science*, 325(5941):698–701, 2009.

[66] S. Balachandar and J. K. Eaton. Turbulent dispersed multiphase flow. *Annual Review Fluid Mechanics*, 42:111–133, 2010.

[67] K. Agrawal, P. N. Loezos, M. Syamlal, and S. Sundaresan. The role of meso-scale structures in rapid gas-solid flows. *Journal of Fluid Mechanics*, 445:151–186, 2001.

[68] B. J. Glasser, S. Sundaresan, and I. G. Kevrekidis. From bubbles to clusters in fluidized beds. *Physical Review Letters*, 81:1849, 1998.

[69] J. K. Eaton and J. R. Fessler. Preferential concentration of particles by turbulence. *International Journal of Multiphase Flow*, 20:169–209, 1994.

[70] K. Agrawal, W. Holloway, C. C. Milioli, F. E. Milioli, and S. Sundaresan. Filtered models for scalar transport in gas–particle flows. *Chemical Engineering Science*, 95:291–300, 2013.

[71] H. Goyal, O. Desjardins, P. Pepiot, and J. Capecelatro. A computational study of the effects of multiphase dynamics in catalytic upgrading of biomass pyrolysis vapor. *AIChE Journal*, 64:3341–3353, 2018.

[72] Z. Wang, J. Fan, K. Luo, and K. Cen. Immersed boundary method for the simulation of flows with heat transfer. *International Journal of Heat and Mass Transfer*, 52(19):4510–4518, 2009.

[73] Z. G. Feng and E. E. Michaelides. Heat transfer in particulate flows with direct numerical simulation (DNS). *International Journal of Heat and Mass Transfer*, 52(3):777–786, 2009.

[74] N. G. Deen, S. H. L. Kriebitzsch, M. A. van der Hoef, and J. A. M. Kuipers. Direct numerical simulation of flow and heat transfer in dense fluid–particle systems. *Chemical Engineering Science*, 81:329–344, 2012.

[75] S. Tenneti, B. Sun, R. Garg, and S. Subramaniam. Role of fluid heating in dense gas–solid flow as revealed by particle-resolved direct numerical simulation. *International Journal of Heat and Mass Transfer*, 58:471–479, 2013.

[76] B. Sun, S. Tenneti, and S. Subramaniam. Modeling average gas–solid heat transfer using particle-resolved direct numerical simulation. *International Journal of Heat and Mass Transfer*, 86:898–913, 2015.

[77] Q. Xiong and S.-C. Kong. High-resolution particle-scale simulation of biomass pyrolysis. *ACS Sustainable Chemistry & Engineering*, 4(10):5456–5461, 2016.

[78] Q. Xiong, F. Xu, Y. Pan, Y. Yang, Z. Gao, S. Shu, K. Hong, F. Bertrand, and J. Chaouki. Major trends and roadblocks in CFD-aided process intensification of biomass pyrolysis. *Chemical Engineering and Processing-Process Intensification*, 127:206–212, 2018.

[79] D. Gidaspow. *Multiphase flow and fluidization: continuum and kinetic theory descriptions*. Academic Pr, 1994.

[80] Y. Tsuji, T. Kawaguchi, and T. Tanaka. Discrete particle simulation of two-dimensional fluidized bed. *Powder technology*, 77(1):79–87, 1993.

[81] J. Capecelatro and O. Desjardins. An Euler–Lagrange strategy for simulating particle-laden flows. *Journal of Computational Physics*, 238:1–31, 2013.

[82] Q. Xiong, S.K. Kong, and A. Passalacqua. Development of a generalized numerical framework for simulating biomass fast pyrolysis in fluidized-bed reactors. *Chemical Engineering Science*, 99:305–313, 2013.

[83] Qingluan Xue, TJ Heindel, and R. O. Fox. A CFD model for biomass fast pyrolysis in fluidized-bed reactors. *Chem. Eng. Sci.*, 66(11):2440–2452, 2011.

[84] A.A. Boateng and P.L. Mtui. CFD modeling of space-time evolution of fast pyrolysis products in a bench-scale fluidized-bed reactor. *Applied Thermal Engineering*, 33:190–198, 2012.

[85] O. Desjardins, G. Blanquart, G. Balarac, and H. Pitsch. High order conservative finite difference scheme for variable density low Mach number turbulent flows. *Journal of Computational Physics*, 227(15):7125–7159, 2008.

[86] M.H. Kasbaoui, D.L. Koch, and O. Desjardins. Clustering in Euler–Euler and Euler–Lagrange simulations of unbounded homogeneous particle-laden shear. *Journal of Fluid Mechanics*, 859:174–203, 2019.

[87] J. Bruchmuller, B.G.M. van Wachem, S. Gu, K.H. Luo, and R.C. Brown. Modeling the thermochemical degredation of biomass inside a fast pyrolysis fluidized bed reactor. *AIChE Journal*, 00(0):1–13, 2011.

[88] C. Cheng, J. Werther, S. Heinrich, H. Qi, and E. Hartge. CPFD simulation of circulating fluidized bed risers. *Powder Technology*, 235:238–247, 2013.

[89] A.H. Mahmoudi, M. Markovic, B. Peters, and B. Gerrit. An experimental and numerical study of wood combustion in a fixed bed using euler-lagrange approach (xdem). *Fuel*, 150:573–582, 2015.

[90] R.K. Thapa, A. Frohner, G. Tondl, C. Pfeifer, and B.M. Halvorsen. Circulating fluidized bed combustion reactor: Computational particle fluid dynamic model validation and gas feed position optimization. *Computers and Chemical Engineering*, 92:180–188, 2016.

[91] X. Ku, T. Li, and T. Løvås. Euler-Lagrangian simulation of biomass gasification behavior in a high-temperature entrained-flow reactor. *Energy&Fuels*, 28:5184–5196, 2014.

[92] Q. Xiong, Y. Yang, F. Xu, Y. Pan, J. Zhang, K. Hong, G. Lorenzini, and S. Wang. Overview of computational fluid dynamics simulation of reactor-scale biomass pyrolysis. *ACS Sustainable Chemistry & Engineering*, 5(4):2783–2798, 2017.

[93] S. Wang, K. Luo, C. Hu, L. Sun, and J. Fan. Impact of operating parameters on biomass gasification in a fluidized bed reactor: An Euler-Lagrangian approach. *Powder Technology*, 333:304–316, 2018.

[94] Konstantinos Papadikis, S Gu, and Anthony V Bridgwater. CFD modelling of the fast pyrolysis of biomass in fluidised bed reactors: modelling the impact of biomass shrinkage. *Chem. Eng. J.*, 149(1):417–427, 2009.

[95] R.C. Brown. Biorenewable resources: Engineering new products from agriculture. *Iowa State Press: Ames, IA*, 2003.

[96] D. Lathouwers and J. Bellan. Modeling of dense gas–solid reactive mixtures applied to biomass pyrolysis in a fluidized bed. *Int. J. Multiphase Flow*, 27(12): 2155–2187, 2001.

[97] E.W. Lemmon and R.T. Jacobsen. Viscosity and thermal conductivity equations for nitrogen, oxygen, argon and air. *International Journal of Thermophysics*, 25(1):21–69, 2004.

[98] J. Capecelatro, O. Desjardins, and R. O. Fox. Effect of domain size on fluid-particle statistics in homogeneous, gravity-driven, cluster-induced turbulence. *Journal of Fluids Engineering*, 138:1–8, 2016.

[99] Y. Igci, A.T. Andrews, S. Sundaresan, S. Pannala, and T. O'Brien. Filtered two-fluid models for fluidized gas-particle suspensions. *AIChE J.*, 54(6):1431–1448, 2008.

[100] A. Ozel, P. Fede, and O. Simonin. Development of filtered Euler-Euler two-phase model for circulating fluidised bed: High resolution simulation, formulation and a priori analyses. *Int. J. Multiphase Flow*, 55:43–63, 2013.

[101] J. Capecelatro and O. Desjardins. Numerical characterization and modeling of particle clustering in wall-bounded vertical risers. *Chemical Engineering Journal*, 245:295–310, 2014.

[102] P. D. Noymer and L. R. Glicksman. Descent velocities of particle clusters at the wall of a circulating fluidized bed. *Chem. Eng. Sci.*, 55(22):5283–5289, 2000.

[103] A.G.W. Bradbury, Y. Sakai, and F. Shafizadeh. A kinetic model for pyrolysis of cellulose. *Journal of Applied Polymer Science*, 23:3271–3280, 1979.

[104] R.S. Miller and J. Bellan. A generalized biomass pyrolysis model based on superimposed cellulose, hemicellulose, and lignin kinetics. *Combustion Science and Technology*, 126:97–138, 1997.

[105] A.G. Liden, F. Berruti, and D.S. Scott. A kinetic model for the production of liquids from the flash pyrolysis of biomass. *Chemical Engineering Communications*, 65, 1988.

[106] T. B. Anderson and R. Jackson. Fluid mechanical description of fluidized beds. Equations of motion. *Industrial & Engineering Chemistry Fundamentals*, 6(4): 527–539, 1967.

[107] L. G. Gibilaro, K. Gallucci, R. Di Felice, and P. Pagliai. On the apparent viscosity of a fluidized bed. *Chemical Engineering Science*, 62(1-2):294–300, 2007.

[108] V. S. Sutkar, N. G. Deen, A. V. Patil, V. Salikov, S. Antonyuk, S. Heinrich, and J. A. M. Kuipers. Chemical engineering journal. *Chemical Engineering Journal*, 288:185–197, 2016.

[109] C. D. Pierce. *Progress-variable approach for large-eddy simulation of turbulent combustion*. PhD thesis, Stanford University, 2001.

[110] H. Herrmann, G. Blanquart, and V. Raman. Flux corrected finite volume scheme for preserving scalar boundedness in reacting large-eddy simulations. *AIAA journal*, 44, 2006.

[111] S. Tenneti and S. Subramaniam. Drag law for monodisperse gas-solid systms using particle-resolved direct numerical simulation of flow past fixed assemblies of spheres. *International Journal of Multiphase Flow*, 37(9):1072–1092, 2011.

[112] P. A. Cundall and O. D. L. Strack. A discrete numerical model for granular assemblies. *Geotechnique*, 29(1):47–65, 1979.

[113] D.J. Gunn. Transfer or heat and mass to particles in fixed and fluidised beds. *International Journal of Heat and Mass Transfer*, 21:467–476, 1978.

[114] J. Capecelatro, O. Desjardins, and R. O. Fox. On fluid–particle dynamics in fully developed cluster-induced turbulence. *Journal Fluid Mechanics*, 780: 578–635, 2015.

[115] N. Peters. Turbulent combustion. *Cambridge University Press*, 2000.

[116] P. Moin, K. Squires, W. Cabot, and S. Lee. A dynamic subgrid-scale model for compressible turbulence and scalar transport. *Phys. Fluids*, 3(11):2746–2757, 1991.

[117] H. Tennekes and J. L. Lumley. *A first course in turbulence*. MIT Press, Cambrige, Mass., 1992.

[118] F. R. Menter. Two-equation eddy-viscosity turbulence models for engineering applications. *AIAA Journal*, 32, 1994.

[119] F. R. Menter, M. Kuntz, and R. Langtry. Ten years of industrial experience with the SST turbulence model. *Turbulence, heat and mass transfer*, 4(625), 2003.

[120] J. Slotnick, A. Khodadoust, J. Alonso, D. Darmofal, W. Gropp, E. Lurie, and D. Mavriplis. FD vision 2030 study: a path to revolutionary computational aerosciences. *Tech. Rep.*, 2014.

[121] R. H. Bush, T. Chyczewski, K. Duraisamy, B. Eisfeld, C. L. Rumsey, and B. R. Smith. Recommendations for future efforts in rans modeling and simulation. *AIAA SchiTech 2019 Forum*, 2019.

[122] M. P. Brenner, J. D. Eldredge, and J. B. Freund. Perspective on machine learning for advancing fluid mechanics. *Physical Review Fluids*, 4(10), 2019.

[123] C. L. Rumsey. Successes and challenges for flow control simulations (invited). *AIAA, 4th flow Control Conference, 23-26 June 2008, Seattle, WA*, 2008.

[124] H. Xiao, J.-L. Wu, S. Laizet, and L. Duan. Flows over periodic hills of parameterized geometries: a dataset for data-driven turbulence modeling from direct simulations. *Computers and Fluids*, 2020.

[125] F. Kohler, J. Munz, and M. Schafer. Data-driven augmentation of rans turbulence models for improved prediction of separation in wall-bounded flows. *AIAA SchiTech 2020 Forum*, 2020.

[126] B. Parmar, E. Peters, K. E. Jansen, A. Doostan, and J. A. Evans. Generalized non-linear eddy viscosity models for data-assisted reynolds stress closure. *AIAA SchiTech 2020 Forum*, 2020.

[127] J. Capecelatro, O. Desjardins, and R. O. Fox. Strongly–coupled gas–particle flows in vertical channels. part ii: Turbulence modeling. *Physics of Fluids*, 28: 1–22, 2016.

[128] J.-X. Wang, J. Wu, and H. Xiao. Physics informed machine learning approach for reconstructing Reynolds stress modeling discrepancies based on dns data. *Physical Review Fluids*, 2(3):1–22, 2017.

[129] R. Tibshirani. Regression shrinkage and selection via the LASSO. *Journal of the Royal Statistical Society B*, 58(1):267–288, 1996.

[130] H. Zou and T. Hastie. Regularization and variable selection via the elastic net. *Journal of the Royal Statistical Society B*, 67(2):301–320, 2005.

[131] C.M. Bishop. *Pattern recognition and machine learning*. Springer Science+Business Media, LLC, 2006.

[132] S. B. Pope. A more general effective-viscosity hypothesis. *Journal of Fluid Mechanics*, 72(2):331–340, 1975.

[133] C. G. Speziale, S. Sarkar, and Gatski T. B. Modelling the pressure-strain correlation of turbulence: an invariant dynamical systems approach. *Journal of Fluid Mechanics*, 227:245–272, 1991.

[134] K. Hanjalic and B. E. Launder. A Reynolds stress model of turbulence and its application to thin shear flows. *Journal of Fluid Mechanics*, 52:609–638, 1972.

[135] B. E. Launder. Phenomenological modeling: present... and future? in J.L. Lumley (Ed.). *Whither Turbulence? Turbulence at the Crossroads*, pages 439–485, 1990.

[136] T. B. Gatski and C. G. Speziale. On explicit algebraic stress models for complex turbulent flows. *Journal of Fluid Mechanics*, 254:59–78, 1993.

[137] M. H. Kasbaoui, R. Patel, D. Koch, and O. Desjardins. An algorithm for solving the Navier–Stokes equations with shear-periodic boundary conditions and its application to homogeneously sheared turbulence. *Journal of Fluid Mechanics*, 833:687–716, 12 2017. doi: 10.1017/jfm.2017.734.

[138] T. Passot and A. Pouquet. Numerical simulation of compressible homogeneous flows in the turbulent regime. *Journal of Fluid Mechanics*, 181:441–466, 1986.

[139] K. R. Sreenivasan. On the universality of the Kolmogorov constant. *Physics of Fluids*, 7:2778–2784, 1995.

[140] J. C. Rotta. Statistische theorie nichthomogener turbulenz. *Z. Phys.*, 129: 547–572, 1951.

[141] C. G. Speziale. Turbulence modeling in noninertial frames of reference. *Theoretical and Computational Fluid Dynamics*, 1:3–19, 1989.

[142] T. Jongen, G. Mompean, and T. B. Gatski. Accounting for Reynolds stress and dissipation rate anisotropies in inertial and noninertial frames. *Physics of Fluids*, 10(674), 1998.

[143] Y. Zhao, H. D. Akolekar, J. Weatheritt, V. Michelassi, and R. D. Sandberg. Turbulence model development using CFD-driven machine learning, 2019.

[144] J. Wu, H. Xiao, R. Sun, and Q. Wang. Reynolds-averaged Navier–Stokes equations with explicit data-driven Reynolds stress closure can be ill-conditioned. *Journal of Fluid Mechanics*, 869:553–586, 2019.

[145] H. G. Weller, G. Tabor, H. Jasak, and C. Fureby. A tensorial approach to computational continuum mechanics using object-oriented techniques. *Computers in Physics*, 12(6), 1998.

[146] B. Krank, M. Kronbichler, and W. A. Wall. Direct numerical simulation of flow over periodic hills up to Re_H =10,595. *Flow Turbulence and Combustion*, 101: 521–551, 2018.

[147] N. Kasagi and A. Matsunaga. Three-dimensional particle-tracking velocimetry measurement of turbulence statistics and energy budget in a backward-facing step flow. *International Journal of Heat and Fluid Flow*, 16(6):477–485, 1995.

[148] M. A. Hopkins and M. Y. Louge. Inelastic microstructure in rapid granular flows of smooth disks. *Physics of Fluids*, 3(1):47–57, 1991.

[149] I. Goldhirsch and G. Zanetti. Clustering instability in dissipative gases. *Physical Review Letters*, 70(11):1619, 1993.

[150] J. J. Wylie and D. L. Koch. Particle clustering due to hydrodynamic interactions. *Physics of Fluids*, 12(5):964–970, 2000.

[151] D. C. Miller, M. Syamlal, D. S. Mebane, C. Storlie, D. Bhattacharyya, N. V. Sahinidis, D. Agarwal, C. Tong, S. E. Zitney, A. Sarkar, X. Sun, S. Sundaresan, E. Ryan, D. Engel, and C. Dale. Carbon capture simulation initiative: a case study in multiscale modeling and new challenges. *Annual Review of Chemical and Biomolecular Engineering*, 5:301–323, 2014.

[152] H. Pouransari and A. Mani. Effects of preferential concentration on heat transfer in particle- based solar receivers. *Journal of Solar Energy Engineering*, 139: 021008, 2017.

[153] D. Besnard, F. H. Harlow, and R. M. Rauenzhan. Conservation and transport properties of turbulence with large density variations. Report LA-10911-MS, 1987.

[154] R. Rzehak and E. Krepper. Cfd modeling of bubble-induced turbulence. *International Journal of Multiphase Flow*, 55:138–155, 2013.

[155] T. Ma, D. Lucas, and A. D. Bragg. Explicit algebraic relation for calculating Reynolds normal stresses in flows dominated by bubble-induced turbulence. *Physical Review Fluids*, 5:084305, 2020.

[156] L. P. Wang and M. R. Maxey. Settling velocity and concentration distribution of heavy particles in homogeneous isotropic turbulence. *Journal of Fluid Mechanics*, 256:27–68, 1993.

[157] A. Aliseda, A. Cartellier, F. Hainaux, and J. C. Lasheras. Effect of preferential concentration on the settling velocity of heavy particles in homogeneous isotropic turbulence. *Journal of Fluid Mechanics*, 468:77–105, 2002.

[158] T. S. Yang and S. S. Shy. The settling velocity of heavy particles in an aqueous near-isotropic turbulence. *Physics of Fluids*, 15(4):868–880, 2003.

[159] A. Ferrante and S. Elghobashi. On the physical mechanisms of two-way coupling in particle-laden isotropic turbulence. *Physics of Fluids*, 15(2):315–329, 2003.

[160] T. Bosse, L. Kleiser, and E. Meiburg. Small particles in homogeneous turbulence: Settling velocity enhancement by two-way coupling. *Physics of Fluids*, 18(2):027102, 2006.

[161] G. K. Batchelor. Sedimentation in a dilute dispersion of spheres. *Journal of Fluid Mechanics*, 52:245–268, 1972.

[162] G. K. Batchelor. Sedimentation in a dilute polydisperse system of interacting spheres. Part 1. General theory. *Journal of Fluid Mechanics*, 119:379–408, 1982.

[163] G. K. Batchelor. A new theory of the instability of a uniform fluidized bed. *Journal of Fluid Mechanics*, 193:75–110, 1988.

[164] J. Capecelatro, O. Desjardins, and R. O. Fox. On the transition between turbulence regimes in particle-laden channel flows. *Journal of Fluid Mechanics*, 845:499, 2018.

[165] B. M. Johnson and O. Schilling. Reynolds-averaged Navier–Stokes model predictions of linear instability. I: Buoyancy-and shear-driven flows. *Journal of Turbulence*, 12:N36, 2011.

[166] Y. Zhou. Rayleigh–Taylor and Richtmyer–Meshkov instability induced flow, turbulence, and mixing. II. *Physics Reports*, 723-725:1–160, 2017.

[167] R. F. Mudde. Gravity-driven bubbly flows. *Annual Review of Fluid Mechanics*, 37:393–423, 2005.

[168] M. I. Jordan and T. M. Mitchell. Machine learning: Trends, perspectives, and prospects. *Science*, 349(6245):255–260, 2015.

[169] K. Duraisamy, Z. J. Zhang, and A. P. Singh. New approaches in turbulence and transition modeling using data-driven techniques. *53rd AIAA Aerospace Sciences Meeting*, page 1284, 2015.

[170] M. Schmelzer, R. P. Dwight, and P. Cinnella. Discovery of algebraic Reynolds-stress models using sparse symbolic regression. *Flow, Turbulence and Combustion*, 104(2):579–603, 2020.

[171] Z. Sun and J. Zhu. A consolidated flow regime map of upward gas fluidization. *AIChE Journal*, 65:1–15, 2019.

[172] Shivshankar Sundaram and Lance R Collins. A numerical study of the modulation of isotropic turbulence by suspended particles. *Journal of Fluid Mechanics*, 379:105–143, 1999.

[173] T. Hastie, R. Tibshirani, and J. Friedman. *The Elements of Statistical Learning: Data Mining, Inference and Prediction*. Springer, New York, NY, 2009. ISBN 978-0-387-84857-0.

[174] B. M. de Silva, K. Champion, M. Quade, J.-C. Loiseau, J. N. Kutz, and S. L. Brunton. Pysindy: A python package for the sparse identification of nonlinear dynamics from data. 2020.

[175] E. M. Sparrow, T. M. Hallman, and R. Siegel. Turbulent heat transfer in the thermal entrance region of a pipe with uniform heat flux. *Applied Scientific Research, Section A*, 7:37–52, 1957.

[176] P. H. Abbrecht and S. W. Churchill. The thermal entrance region in fully developed turbulent flow. *AIChE Journal*, 6:268–273, 1960.

[177] Y. Lee. Turbulent heat transfer from the core tube in thermal entrance region of concentric annuli. *International Journal of Heat and Mass Transfer*, 11: 509–522, 1968.

[178] B. O. Hasan. Heat transfer analysis in thermal entrance region under turbulent flow conditions. *Asia-Pacific Journal of Chemical Engineering*, 8:578–592, 2012.

[179] Frank P. Incropera. *Fundamentals of heat and mass transfer*. Wiley, 2011.

[180] James R. Welty, Gregory L. Rorrer, and David G. Foster. *Fundamentals of momentum, heat, and mass transfer*. John Wiley & Sons, Inc., 2019.

[181] Z. Xu and Y. Zhang. Quench rates in air, water, and liquid nitrogen, and interference of temperature in volcanic eruption columns. *Earth and Planetary Science Letters*, 200:315–330, 2002.

[182] L. Wilson, R. S. J. Sparks, T. C. Huang, and N. D. Watkins. The control of volcanic column heights by eruption energetics and dynamics. *Journal of Geophysical Research*, 83:1829–1836, 1978.

[183] G. Lube, E. C. P. Breard, T. Esposti-Ongaro, J. Dufek, and B. Brand. Multiphase flow behaviour and hazard prediction ofpyroclastic density currents. *Nature Reviews Earth and Environment*, 1:348–365, 2020.

[184] A. B. Morris, Z. Ma, S. Pannala, , and C. M. Hrenya. Simulations of heat transfer to solid particles flowing through an array of heated tubes. *Solar Energy*, 130:101–115, 2016.

[185] M. Mehos, C. Turchi, J. Vidal, M. Wagner, Z. Ma, C. Ho, W. Kolb, C. Andraka, and A. Kruizenga. Concentrating solar power gen3 demonstration roadmap. *Technical report, National Renewable Energy Laboratory (NREL), Golden, CO, United States*, 2017.

[186] K. Pielichowska and K. Pielichowski. Phase change materials for thermal energy storage. *Progreses in Materials Science*, 65:67–123, 2014.

[187] M. Louge, J. M. Yusof, and J. T. Jenkins. Heat transfer in the pneumatic transport of massive particles. *International Journal of Heat and Mass Transfer*, 36:265–275, 1993.

[188] T. A. Ebert and L. R. Glicksman. Determination of particle and gas convective heat trasnfer components in a circulating fluidized bed. *Chemical Engineering Science*, 48:2179–2188, 1993.

[189] P. D. Noymer and L. R. Glicksman. Cluster motion and particle-convective heat trasnfer at the wall of a circulating fluidized bed. *International Journal of Heat and Mass Transfer*, 41:147–158, 1998.

[190] H. Lei, L.-T. Zhu, and Z.-H. Luo. Study of filtere interphase heat transfer using highly resolved CFD-DEM simulations. *AIChE Journal: Particle Techology and Fluidization*, pages 1–12, 2020.

[191] S. Rauchenzauner and S. Schneiderbauer. A dynamic spatially averaged two-fluid model for heat transport in moderately dense gas–particle flows. *Physics of Fluids*, 32:063307:1–20, 2020.

[192] L. Jofre, Z.R. del Rosario, and G. Iaccarino. Data-driven dimensional analysis of heat transfer in irradiated particle-laden turbulent flow. *International Journal of Multiphase Flow*, 125:1–15, 2020.

[193] C. Peng, B. Kong, J. Zhou, B. Sun, A. Passalacqua, S. Subramaniam, and R. O. Fox. Implementation of pseudo-turbulence closures in an Eulerian–Eulerian two-fluid model for non-isothermal gas–solid flow. *Chemical Engineering Science*, 207:663–671, 2019.

[194] A. Yousefi, M.N. Ardekani, F. Picano, and L. Brandt. Regimes of heat transfer in finite-size particle suspensions. *International Journal of Heat and Mass Transfer*, 177:121514, 2021.

[195] M. T. Shah, R. P. Utikar, V. K. Pareek, G. M. Evans, and J. B. Joshi. Computational fluid dynamic modelling of FCC riser: A review. *Chemical Engineering Research and Design*, 111:403–448, 2016.

[196] I. N. Hamdhan ad B. G. Clarke. Determination of thermal conductivity of coarse and fine sand soils. *Proceedings World Geothermal Congress 2010*, Bali, Indonesia, 25-29 April 2010, 2010.

[197] L. Qiu, V. Murashov, and M. A. White. Zeolite 4A: heat capacity and thermodynamics properties. *Solid State Sciences*, 2:841–846, 2000.

[198] A. S. Issangya, J.R Grace, D. Bai, and J. Zhu. Further measurements of flow dynamics in a high-density circulating fluidized bed riser. *Powder Technology*, 111:104–113, 2000.

[199] D. Searson. Gptips: Genetic programming & symbolic regression for MATLAB. *http://gptips.sourceforge.net*, 2009.

[200] Maximilian Reissmann, Josef Hasslberger, Richard D Sandberg, and Markus Klein. Application of Gene Expression Programming to a-posteriori LES modeling of a Taylor Green vortex. *Journal of Computational Physics*, 424:109859, 2021.

[201] Joachim Dominique, Julien Christophe, Christophe Schram, and Richard D Sandberg. Inferring empirical wall pressure spectral models with Gene Expression Programming. *Journal of Sound and Vibration*, 506:116162, 2021.

[202] Saeed Samadianfard. Gene expression programming analysis of implicit colebrook–white equation in turbulent flow friction factor calculation. *Journal of Petroleum Science and Engineering*, 92:48–55, 2012.

[203] Jack Weatheritt and Richard Sandberg. A novel evolutionary algorithm applied to algebraic modifications of the RANS stress–strain relationship. *Journal of Computational Physics*, 325:22–37, 2016.

[204] Yaomin Zhao, Harshal D Akolekar, Jack Weatheritt, Vittorio Michelassi, and Richard D Sandberg. RANS turbulence model development using cfd-driven machine learning. *Journal of Computational Physics*, 411:109413, 2020.